WS No. 19 Mark III

PULSE-WIDTH MODULATION

Note: This information is provisional and is supplied for guidance pending the issue of more complete instructions. All errors of a technical nature should be notified in accordance with Tels. A 009.

CONTENTS

LIST OF FIGURES

GENERAL

Glossary

1. A number of terms now in general use in describing the main features of voltage and current pulses have been lerived from the geometrical characteristics displayed by such pulses when plotted on the usual voltage- or current-time graphs. Since these terms will be used frequently throughout the following discussion of pulse-width modulation, the following explanations are given to avoid confusion. (Reference is made to typical (voltage) pulses illustrated in Fig. 1):—

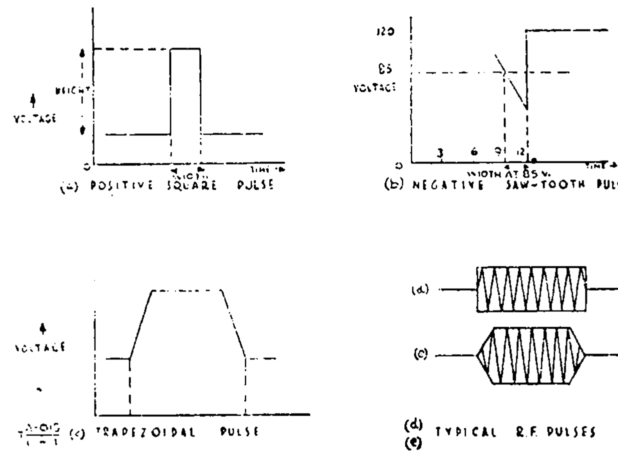


Fig. 1—Types of voltage pulse

(*a*) *D.C. Pulses.*—Voltage or current pulses in which the voltage or current excursion may be considered as being entirely positive or negative with respect to the steady-state conditions of the circuit. Such pulses are often designated "positive" (or "positive-going") and "negative" (or "negative-going") (see Fig. 1 (*a*), (*b*) and (*c*)).

(*b*) *R.F. Pulses.*—Pulses of radio-frequency oscillation whose envelopes have a square or trapezoidal form (see Fig. 1 (*d*) and (*e*)).

(*c*) *Pulse width.*—The duration of a pulse, from onset to completion, usually measured in microseconds. In some cases the width of a pulse at a specified voltage level may be referred to. Thus the width of the pulse in Fig. 1 (*b*) is 3μsec. at 85V.

(*d*) *Leading edge, front edge or front.*—That part of a pulse traced from onset to peak. *A steep-fronted pulse* is one in which the peak is reached in a comparatively short time after the onset.

(*e*) *Trailing edge, back edge or back.*—That part of the pulse traced from peak to completion.

(*f*) *Square (or rectangular) pulse.*—A pulse in which the leading and trailing edges are extremely steep, and the top is flat (see Fig. 1 (*a*)).

(*g*) *Sawtooth pulse.*—A pulse of triangular form in which the leading and trailing edges are (usually) linear, the trailing edge generally being very much steeper than the leading edge (see Fig. 1 (*b*)).

(*h*) *Height or amplitude.*—In the case of a D.C. pulse, this measures the total voltage or current excursion from onset to peak. This is shown in Fig. 1 (*a*), etc. In the case of R.F. pulses, the term amplitude has the usual significance.

(*j*) *Trapezoidal pulse.*—A pulse having a flat top and sloping edges, as shown in Fig. 1 (*c*).

(*k*) *Integration and differentiation.*—A square voltage pulse applied to a resistance-capacity combination of suitable dimensions will develop voltage pulses across the two components whose general forms will be as shown in Fig. 2. The voltage across the condenser is known as the integrated form, and that across the resistance as the differentiated form, of the applied voltage pulse.

(*l*) *Pulse clipping.*—The process of selecting for amplification only that part of a pulse which lies between the peak, or the base, and a specified voltage level (see Fig. 3 (*a*)).

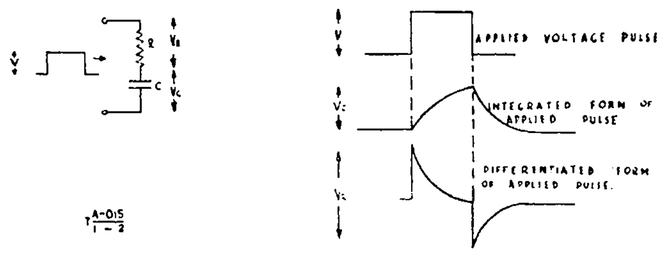


Fig. 2—
Integration and differentiation of a square pulse

(*m*) *Pulse slicing.*—The process of selecting, for amplification, only that part of a pulse which lies between specified voltage levels about the centre of the pulse (see Fig. 3 (*b*)).

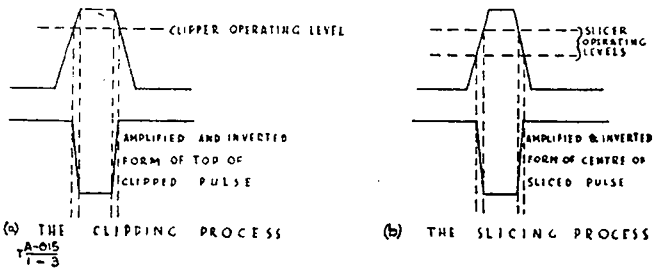


Fig. 3—Pulse clipping and slicing

Modulation—general

2. In any communication system employing modulation the essentials are the "intelligence" and the "carrier." The intelligence will usually have a highly complex wave form which, however, may be analysed into the simple wave forms of all the frequencies present., with their appropriate amplitude and phase relationships. The wave form of the carrier, in the unmodulated condition, will have its three main attributes, amplitude, frequency and phase, all constant. The carrier frequency is usually considerably higher than that of any of the component frequencies of the intelligence. The process of modulation involves impressing upon the carrier the characteristics of the intelligence wave form in such a way that *one* of the main attributes of the carrier wave form is varied linearly in accordance with the configuration of this wave form. It will be convenient, in the first instance, to draw an analogy between normal methods of modulation of a sinusoidal carrier by a *single* intelligence frequency, and modulation, in the various ways possible, of a pulse wave form by the same frequency; and afterwards to proceed to develop the idea of *pulse-width* modulation for simple and more complex intelligence wave forms.

3. Consider first the results of amplitude modulation, and of frequency modulation, of a sinusoidal carrier by modulating wave forms representing, firstly a low, loud note, and secondly a high, soft note, each of a single frequency. (*Note*: The diagrams given in Fig. 4 imply that loudness is a function of amplitude only, and not of frequency. In point of fact, the loudness-frequency characteristic of the human ear for sounds of equal power is complex. For the sake of simplicity, however, the response is taken to be linear for the two frequencies illustrated.) In amplitude modulation, the *extent* of the amplitude variation of the carrier from that in the unmodulated condition measures the comparative *loudness*

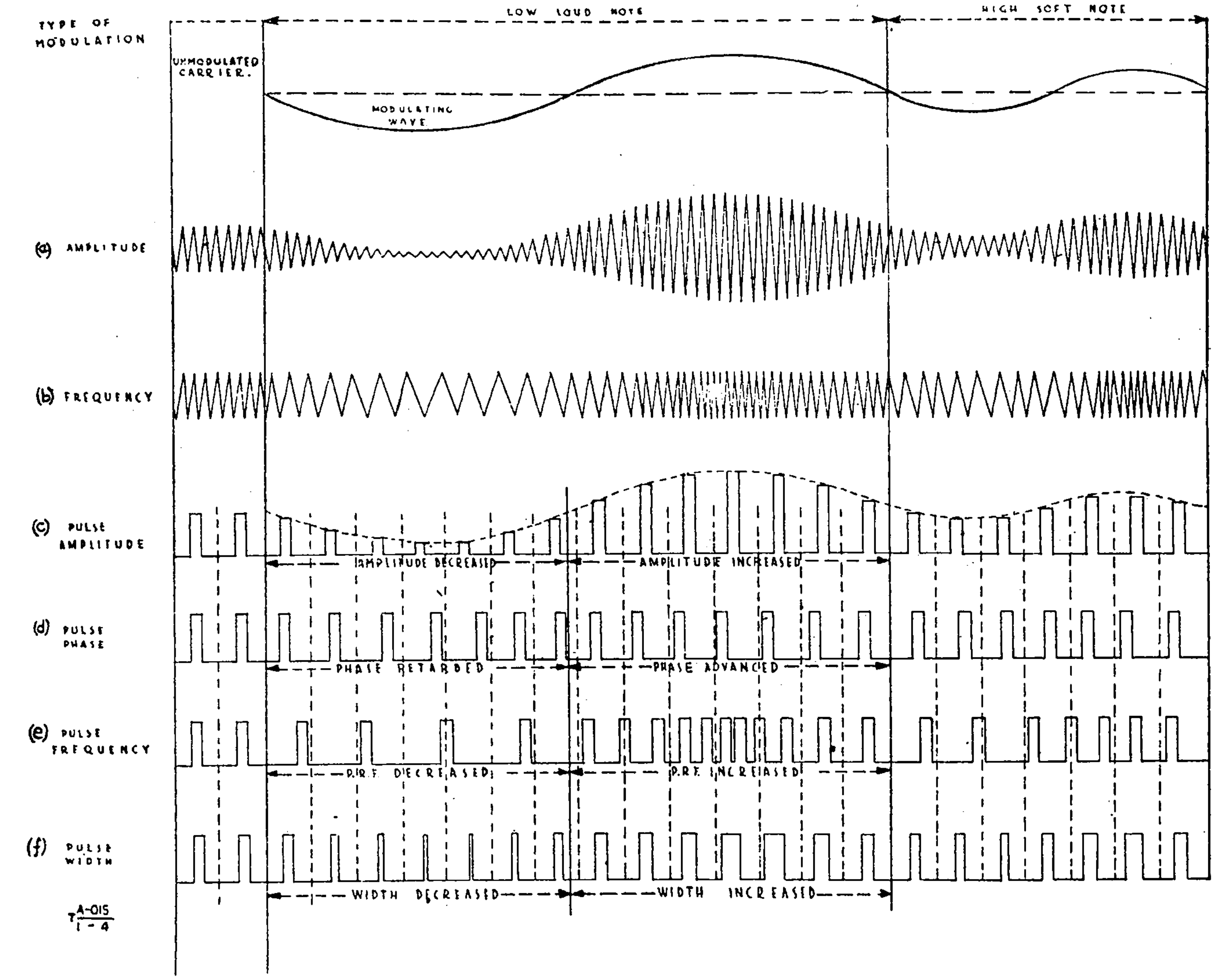


Fig. 4—Diagrammatic representation of different types of modulation

of the note, while the *number* of complete variations in amplitude *per second* represents the *frequency* of the note. This sentence is equally true for the case of frequency modulation if throughout the word frequency be substituted for the word amplitude. The wave forms of the modulating frequency, the unmodulated carrier and the modulated carrier shown in Fig. 4 (*a*) and (*b*) clearly indicate the results of the processes of modulation of amplitude and frequency.

MODULATION OF PULSE WAVE FORMS

4. The term "unmodulated carrier" usually implies a sinusoidal wave form of frequency very much higher that that of any frequency in the intelligence spectrum. The question of the necessity of this comparatively high carrier frequency will be referred to later. It is the carrier wave form that is of immediate interest. Provided that this wave form follows a regular and continuously repeating pattern, and provided that one of its main characteristics is capable of varying linearly in accordance with the modulating wave form, then linear modulation is possible. Thus, for example, let us suppose that the unmodulated carrier consists of a train of equal D.C. pulses, rectangular in form and recurring at equal intervals of time. Such a train of pulses is shown in Fig. 4 (*c*) to (*f*), and, for convenience in reference, a series of lines are drawn vertically on the diagram at equal intervals of time, the pulses in the unmodulated condition being assumed to have constant width and amplitude and to occur regularly at fixed intervals of time after the reference times. The main characteristics of the pulse train which may be varied by the modulating process are :—

(*a*) Amplitude.
(*b*) Phase.
(*c*) Recurrence frequency.
(*d*) Width.

5. The results of modulating the pulse train by varying in turn *one* of these characteristics with a wave form corresponding to a comparatively low, loud note, and a high, soft note, are shown in Fig. 4 (*c*) to (*f*) and should be compared with the corresponding results for a sinusoidal carrier. As in the case of modulation of a sinusoid, so for the modulation of a pulse train, it is the *extent* of the variation in the characteristic of the pulse train that is determined by the *loudness* of the note, while it is the *frequency* with which this variation is carried out that is

determined by the *frequency* of the modulating note. Since all these modes of pulse-train modulation, and others derived from them have been, or are being, employed in experimental or production equipment, it is worth while spending a short time in considering amplitude, phase and recurrence frequency modulation of pulses before proceeding to the more detailed consideration of width modulation.

6. In any form of pulse-train modulation, the relative discontinuity of the unmodulated wave form leads to the necessity of selecting appropriate voltages from the modulating wave form to bring about the modulation of each individual pulse. The voltages appropriate to these pulses will be termed "instantaneous modulating voltages" throughout the following discussion, and, with respect to the mean modulating voltage, may be either positive or negative, the change of sign taking place regularly for a sinusoidal modulating wave.

Pulse-amplitude modulation (P.A.M.)

7. In this mode the heights of individual pulses in the train are so affected by the modulating process that each individual pulse has a final height corresponding to its unmodulated height, with an amount corresponding to the appropriate instantaneous modulating voltage (taken at the onset of the pulse) added to it algebraically. As a direct result, the upper envelope of the modulated pulse train follows the form of the modulating wave, and the process is analogous to amplitude modulation of a sinusoidal carrier (see Fig. 4 (*c*)). The *extent* of the amplitude variation is determined by the *loudness* of the note, while the *frequency* with which the variation is executed is determined by the *frequency* of the note.

Pulse-phase modulation (P.P.M.)

8. In this mode the instants of production of individual pulses are delayed or advanced from the time positions they would occupy in the unmodulated condition by amounts, and in senses, which are dependent upon the values and signs of the corresponding instantaneous modulating voltages in the modulating wave. It might appear from purely geometrical considerations based on Fig. 4 that these instantaneous modulating voltages should be selected from the modulating wave form at time intervals equal to those between the pulses in the unmodulated train, and at instants coincident with, or slightly in advance of, the earliest possible time of formation of the pulses; and that, if the relative displacements of the individual pulses are then made proportional to the appropriate modulating voltages, linear phase modulation will result. This (theoretical) method of deriving the wave form of the phase-modulated pulse train clearly envisages some such initial delay in pulse production, since otherwise the formation of a pulse advanced in phase might have to anticipate the instantaneous modulating voltage controlling its formation. In practice, this method of spacing out the phase-controlling voltages regularly in the modulating wave form is not adopted. Neither, as is shown in paras. 77 to 85, is such a process necessary to secure linear phase modulation. The actual method used (for both pulse-phase, and pulse-width modulation) is discussed from a circuit point of view in paras. 47 to 53, from a mathematical point of view in paras. 75 to 84, and an outline is given in para. 14.

9. For the sake of preserving continuity of ideas, pulse-phase modulation has been dealt with before pulse-width modulation. It should be stated here, however, that a phase-modulated pulse train may be derived directly from a width-modulated train by electrical differentiation, followed by selection of the differentiated pulses corresponding to those edges of the width-modulated train which carry the intelligence. The results of pulse-phase modulation are shown in Fig. 4 (*d*). The *extent* of phase variation of the *pulses* will depend upon the *amplitude* of the modulating wave form, while the *frequency* with which the whole cycle is carried out will be determined by the *frequency* of the modulating wave.

Pulse-recurrence frequency modulation (P.F.M.)

10. Here the recurrence frequency of the pulse train is made to vary to an *extent* dependent upon the *amplitude* of the modulating wave form, and with a *frequency* equal to the *frequency* of this wave. The process is thu analogous to frequency modulation of a sinusoidal carrier, and the results of modulation are shown in Fig. 4 (*e*).

Pulse-width modulation (P.W.M.)

11. In this method of modulation the widths of the pulses are so controlled by the modulating wave form that the width of each individual pulse is varied to an extent, and in a sense, dependent upon the value and the sign of the appropriate instantaneous modulating voltage. Since variation in the width of a pulse will involve displacement of one or both edges along the time axis, i.e., *phase modulation of the pulse edges*, the instants of selection of the modulating voltages controlling the widths of individual pulses will be subject to the same considerations as those encountered for phase modulation of the pulse train, and the same references are therefore made. For the moment, it is sufficient to obtain a general picture so that we may compare all the basic methods of modulation. The process of width modulation will therefore be summarized by saying that to the width of each pulse in the train is added (algebraically) a quantity proportional to the value of the corresponding instantaneous modulating voltage, with the result that linear pulse-width modulation is produced. The *extent* of the width variation of the pulses is determined by the *amplitude*, and the *frequency* with which the variation occurs by the *frequency*, of the modulating wave.

12. To summarize the results of linear modulation by all processes so far discussed, a single graph may now be drawn. The abscissae on this graph will represent the same time scale as that used in Fig. 4, while the ordinates will represent the instantaneous amplitude or instantaneous frequency (for amplitude- and frequency-modulation of a sinusoidal carrier), or instantaneous amplitude, phase, recurrence frequency, or width (for pulse-train modulation), as the case may be. In every case the curve traced by the ordinates will follow the same form as that of the modulating wave. Such a graph is shown in Fig. 5.

13. A number of derived methods of pulse modulation dependent upon the above main methods are possible and in some cases are being employed. These in general involve using a group of pulses to replace the individual pulses in the train discussed above, but full discussion of all possible methods is beyond the scope of this treatment. The main methods in use are tabulated in Fig. 6.

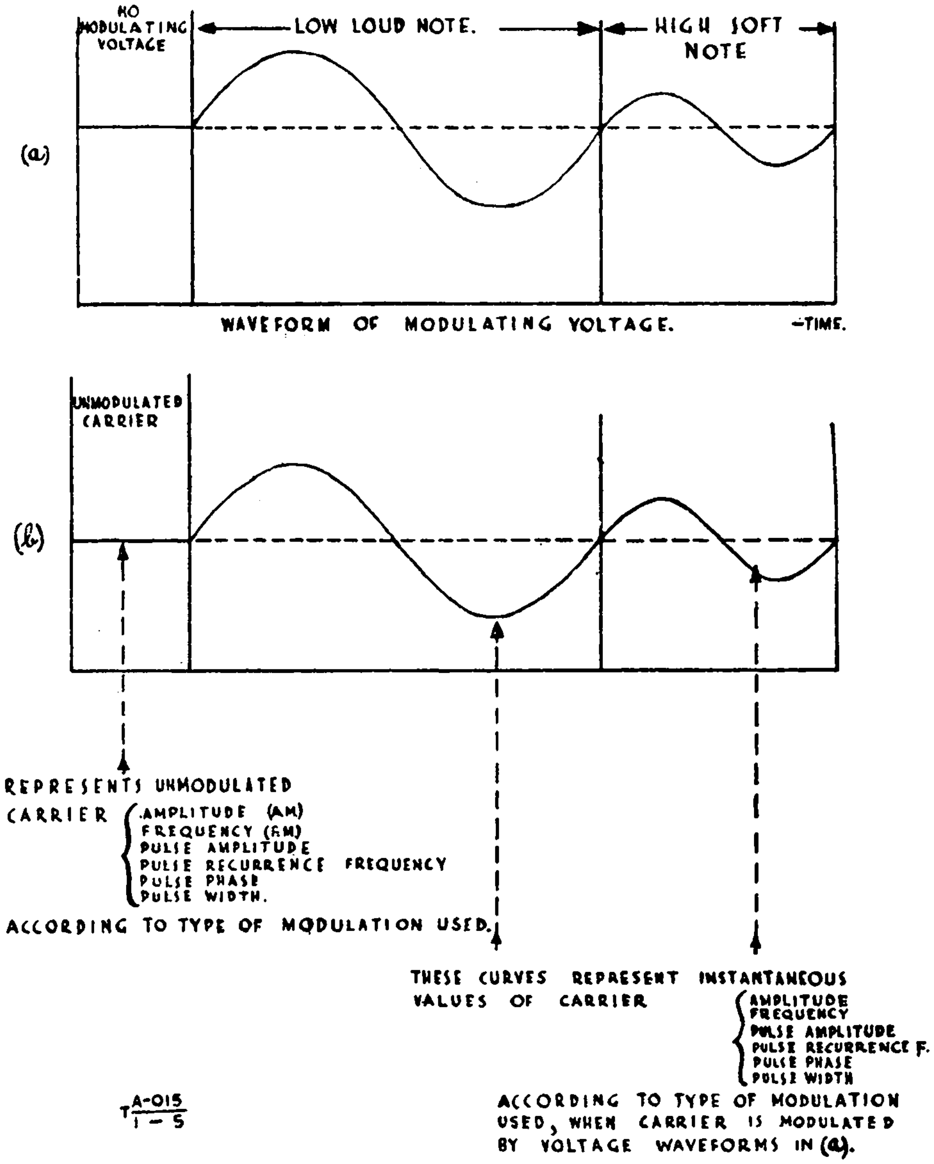


Fig. 5—Linear modulation of a carrier

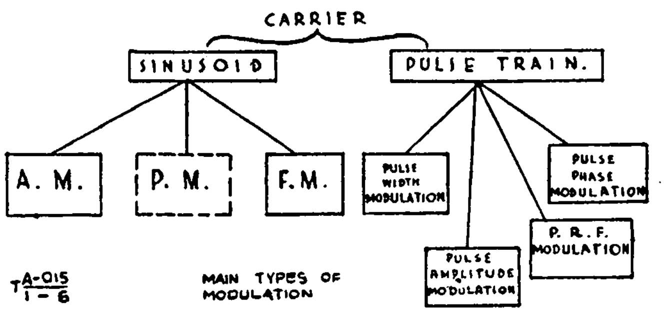


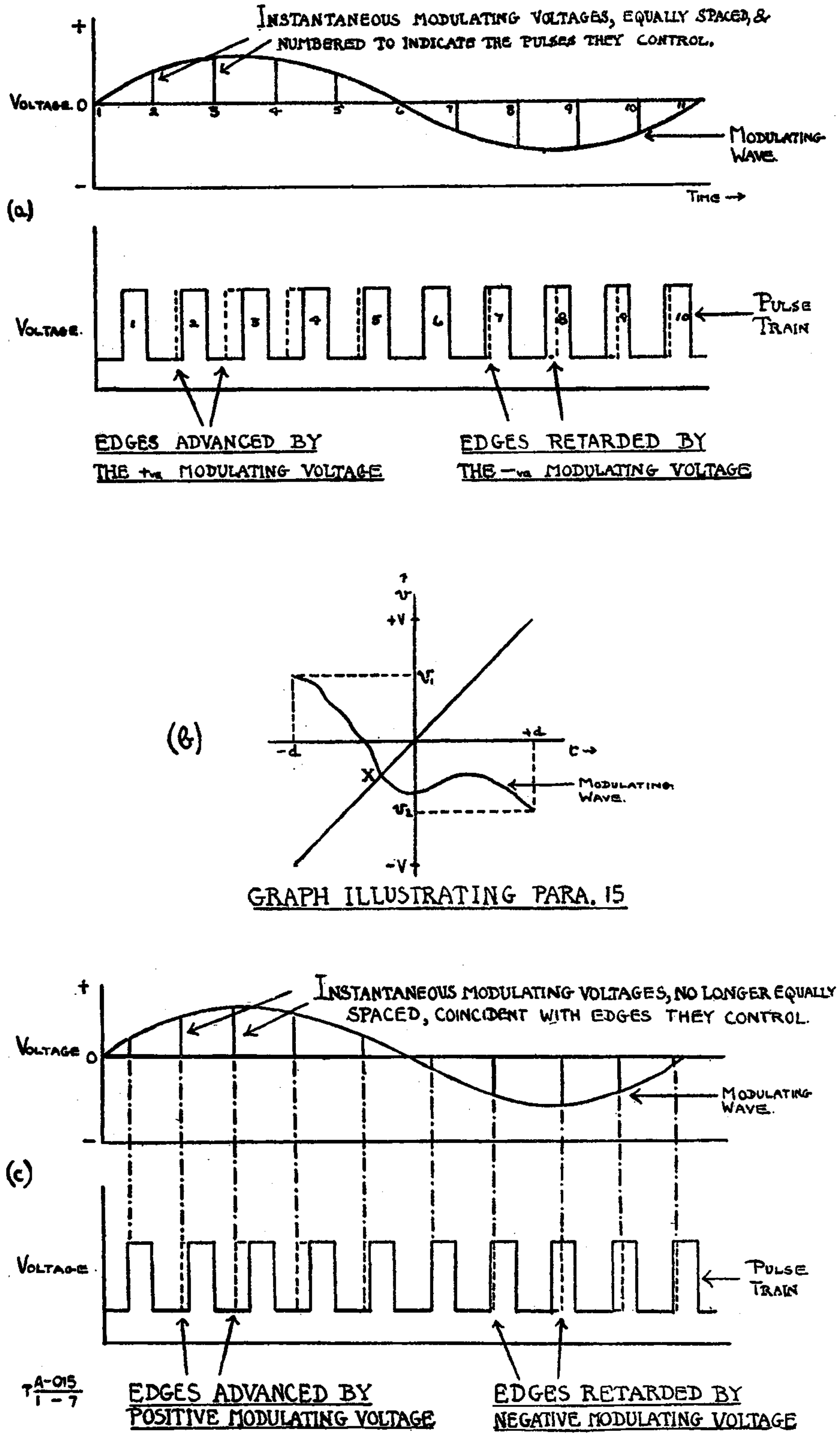
Fig. 6—Main types of modulation

PULSE-WIDTH MODULATION

The particular method of width modulation

14. In Fig. 7 (*a*) a modulating voltage wave form and a ..ain of pulses to be width-modulated by it are shown. ..ow it is possible to bring about variation in the widths of the pulses in three main ways. The leading edges of the pulses may be advanced and retarded, the trailing edges may be advanced and retarded, or both edges may undergo simultaneous displacements in opposite senses by equal amounts, controlled by the modulating wave. In paras. 8 and 11, both of which deal with methods of

pulse modulation involving displacement of one or both pulse edges along the time axis, it is suggested that purely geometrical considerations lead one to expect that for linear modulation the instantaneous modulating voltages controlling the edge displacement of individual pulses should be selected from the modulating wave form at intervals of time equal to those between the unmodulated pulses, and at instants coincident with, or slightly in advance of, the earliest possible moment of formation of the pulses concerned. In Fig. 7 (*a*) such voltages fulfilling these conditions are shown extracted from the modulating wave; front-edge-width moduation is con-



T A-015 / 1-7

Fig. 7—Method of control of pulse width by modulation voltage

sidered, and the dotted lines adjacent to the firm pulse edges indicate the appropriate positions taken up by the edges of the pulses after modulation. These positions have been arrived at by advancing or retarding (according to the sign of the instantaneous modulating voltage) the leading edges of each pulse by an amount proportional to the controlling voltage. The resultant pulse train is in fact linearly modulated by the modulating wave form.

15. This method of modulation and the results obtained from it are based on purely theoretical considerations, and are not encountered in practical devices for producing either width- or phase-modulation of pulse trains. Paras. 77 to 85, dealing with the mathematical aspect of the subject, show that it is not necessary to space out the edge-controlling voltages regularly in the modulating wave form, provided that a particular form of control over the formation of the edges concerned is exerted by the modulating wave. It is this form of control which is employed in practical circuits and the fundamentals of it will now be described. It will be supposed that in the system employed the mean pulse width is a and the maximum allowable deviation from this width is d. Front-edge modulation is considered, since this type of width modulation is employed in practice. It will also be supposed that linear modulation occurs, i.e., that a steady modulating voltage gives rise to a proportional steady change in pulse width. Thus, if the instant at which a given pulse would start in the unmodulated condition is taken as $t = 0$, we may graph against t the modulating voltage required to cause the pulse to start at any instant between $t = -d$ and $t = +d$, and the graph will be a straight line as shown in Fig. 7 (*b*). Also, if over-modulation is to be avoided, the instantaneous value, v, of the modulating voltage must be kept within the limits $-V$ to $+V$, where V is the voltage that causes maximum deviation d in the position of the pulse edge. If now, at time $t = -d$, $v = v_1$, and at time $t = +d$, $v = v_2$, with continuous variation of v, the modulating voltage, between these times, then, whatever form the intermediate variation may take, there must be some instant between $t = -d$ and $t = +d$ at which the modulating voltage takes up the appropriate value to cause the pulse to start *at that instant*. This is shown in Fig. 7 (*b*) by the point X, whose co-ordinates represent the time at which the pulse starts, and the voltage controlling the start.

16. If each pulse in the train has its front edge positioned in this manner, it is clear that the voltages controlling the edge formation are no longer spaced equally along the time axis, since they are coincident with the front edges of the pulses they control. It will be shown that practical circuits behave in this way (paras. 47 to 52), and that the resulting width-modulated pulse train contains components representing the modulating wave form, without distortion, subject to certain restrictions that will be indicated (paras. 84, 85). Linear modulation therefore results from the method. Fig. 7 (*c*) shows the results of modulation on these lines, the instantaneous modulating voltages now being coincident with the pulse edges they control, but no longer spaced regularly along the time axis.

Essentials of a pulse-width modulation transmission system

17. It would be possible to pass a train of pulses, after width modulation, over an ordinary wire system, provided that the pulses were not seriously distorted by the attenuation frequency characteristics of the line. However, such a system would offer little, if any, advantage over more conventional systems, either in the matter of privacy, or for any other reason. It is the employment of pulse-width modulation and other forms of pulse modulation over wireless links that offers the greatest possibilities, and this is the role in which pulse modulation is chiefly employed. The main advantages of this type of communication system will appear in the succeeding text, and will be summarized at the end. For the moment, the requirements of a simple pulse-modulated wireless link employing width modulation will be considered.

18. For transmitting it will be required to produce a train of unmodulated D.C. pulses. This is usually achieved by the use of a pulse generator whose recurrence frequency determines that of the pulse train, and a modulator. The pulses themselves are produced in the modulator and the modulator also receives the intelligence in the form of a voltage wave. The output consists of a train of width-modulated D.C. pulses. These are fed to the modulator of an R.F. oscillator, with the production of a train of R.F. pulses, the envelope of which conforms exactly with that of the modulated D.C. pulse train. The final R.F. output is then taken to a suitable aerial system.

19. For reception, a normal type of superhet. receiver will be required to amplify, and detect the envelope of, the incoming R.F. pulse train. The regenerated D.C. pulses thus obtained may, in the simplest system, be passed direct to a telephone or other acoustical or mechanical device capable of reproducing the intelligence, since, as is shown in the mathematical treatment (paras. 77 to 84), the wave form contains all the intelligence frequencies. Where, however, the pulse recurrence frequency itself falls within the audio spectrum, a second detector, consisting essentially of an integrating circuit followed by a low-pass filter, may be used, this arrangement serving to attenuate the pulse recurrence frequency and its harmonics until they are no longer audible. This second detector does, in fact, convert the width variations of the pulses into corresponding voltage variations, so producing a voltage wave form at the output corresponding to that originally used to modulate the pulse train.

20. It is thus clear that a pulse-width-modulated wireless system must employ two processes of modulation and one or two processes of detection. A block schematic of the essentials of such a system is shown in Fig. 8, the components listed providing, of course, for transmission in only one direction. Duplex working will require complete duplication of the equipment. It is not considered necessary in this treatment to deal with the radio-frequency

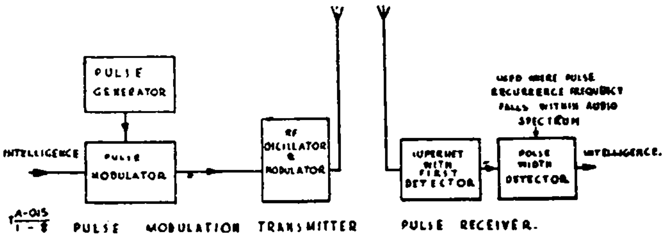


Fig. 8—Essentials of a pulse-width modulation wireless link

transmitter and its modulator in detail, since these follow on conventional and well-established lines. The circuits for producing and shaping the modulated D.C. pulse train are described in paras. 47 to 53. Some extremely important aspects of pulse-width modulation are, however, bound up with the question of receiver design, and of the circuits following the first detector, and these aspects will be considered in the succeeding paragraphs.

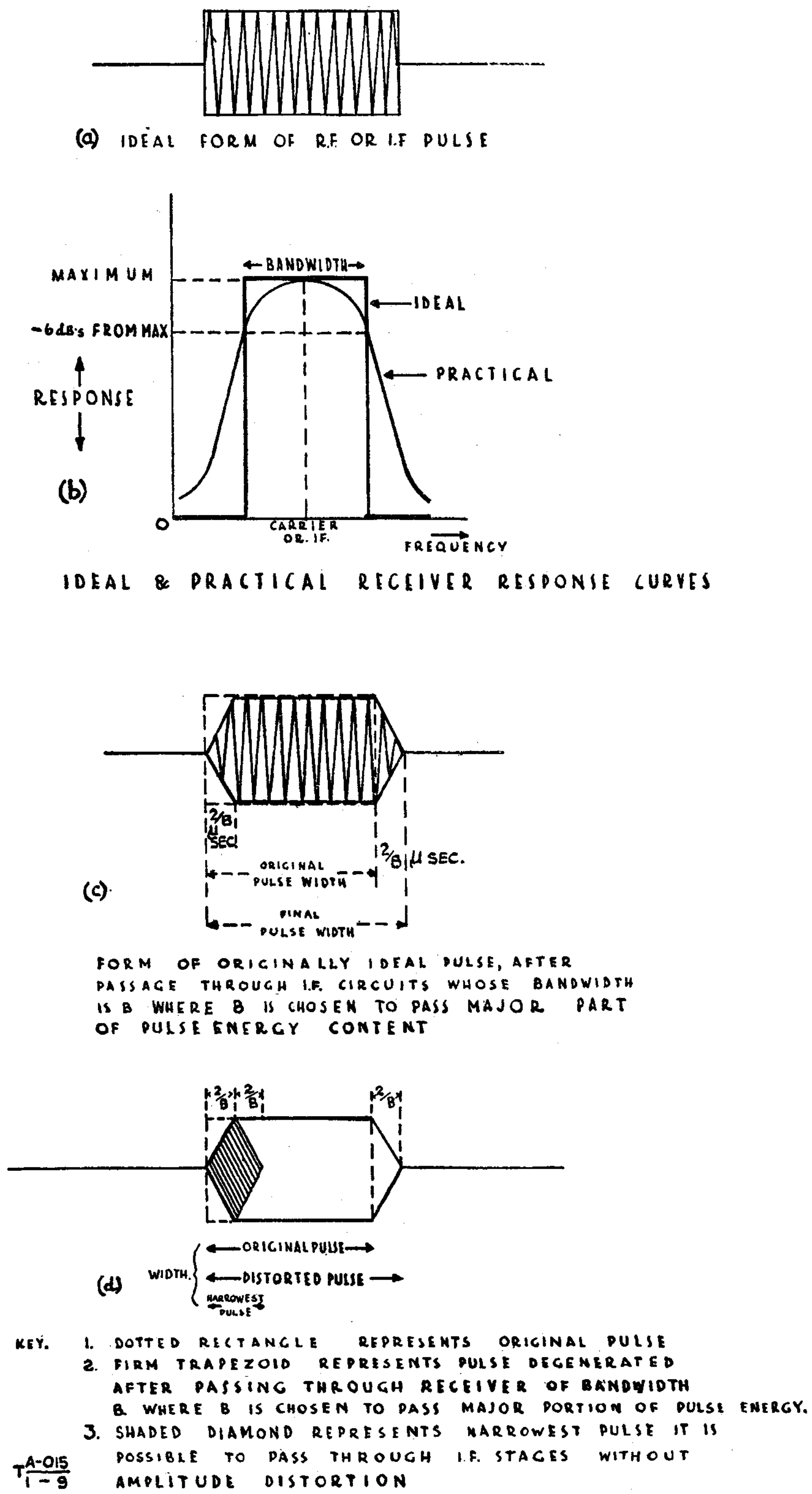


Fig. 9—Pulse distortion produced in passing through I.Fs.

THE P.W.M. RADIO RECEIVER

Band-width requirements

21. In the foregoing statement of the requirements of a pulse-width-modulated wireless transmission system, it was taken for granted that the pulse train, in its R.F. (or I.F.) form, would pass through the receiver without relative attenuation of those component frequencies in the wave form necessary to preserve its geometrical characteristics; and that therefore the first detection process (assuming a perfect detector) would produce a D.C. wave form conforming in every respect to its counterpart at the R.F. modulator. In order to assess the band-width requirements of a receiver capable of passing the pulses without undue distortion, it is necessary to know the component frequencies included in the R.F. and I.F. train and also the relative importance of each component frequency in maintaining the shape of the pulse wave form. An analysis of such a wave form has been undertaken, but it is not included in this treatment. It is possible, nevertheless, to give an introduction to the problem, based on physical concepts, and to quote the results obtained. The ideal R.F. (or I.F.) pulse has the form shown diagrammatically in Fig. 9 (*a*), the envelope of the pulse being perfectly rectangular. The analysis referred to above produces two salient facts :—

(*a*) The sinusoidal components of which the rectangular pulse is composed are infinite in number and cover the complete frequency range to infinity.

(*b*) The amplitude distribution of these components is such that the major part of the energy content of the pulse is carried by a band of frequencies symmetrically disposed about the carrier, and contained within well-defined and comparatively narrow limits.

Further, the actual band-width required to pass this major portion of the energy of the pulse through the receiver can be arrived at in the following extremely simple manner. *Take the reciprocal of the shortest pulse length in microseconds, and multiply by two. The result, in megacycles, is the required band-width.* Thus, for instance, if the unmodulated pulse is of 4μsec. duration, and the depth of modulation employed is 75%, the narrowest pulse will be 1μsec., which will give the required band-width of the receiver as 2 Mc/s.

22. In the analysis from which the results quoted above were derived, it was assumed that the response curve of the receiver concerned was of ideal form, having a perfectly flat top, showing uniform response over the pass band, and vertical sides carrying the response down to zero at the fringes of the band, outside which there was no response. Fig. 9 (*b*) shows such a response curve, and on this curve has been superimposed the type of curve encountered in practice, in which the band-width is taken (according to the usual definition) as the frequency interval between two points at 6db. down from the centre frequency, at which maximum response is obtained. It is interesting to note that the practical curve will give results which conform closely to those which might be expected from the theoretical curve, the absence of a flat top being almost compensated for by having some response outside the fringes of the band.

23. Now if an R.F. or derived I.F. pulse of initially square form be passed through a receiver whose band-width is arrived at by the above method, although the major portion of the energy of the pulse may be successfully transferred, the removal of the higher side frequencies in passage will have an effect upon the final geometrical form of the pulse. It is the higher side frequencies which are mainly concerned with the steepness of the leading and trailing edges of the pulse, and the removal of these brings about degeneration of these edges, with the production of an envelope which is, to a first approximation, trapezoidal. The time taken for the degenerated pulse to rise from its onset to maximum value can be shown to be 2/Bμsec., where B is the band-width in Mc/s. The decay of the trailing edge will occupy an equal time, and so the degenerated pulse is not only altered in form, but also lengthened in time by an amount 2/Bμsec. The effect is illustrated in Fig. 9 (*c*), where the firm dotted rectangle outlines the original pulse, and the trapezoidal form indicates the shape and extent of the pulse after passing through the receiver. The envelopes alone are shown in Fig. 9 (*d*) and in this figure the shaded portion indicates the narrowest pulse that can be passed through the receiver without amplitude distortion, any narrower pulse being prevented from attaining its maximum by the action of the band-pass circuits. It is instructive to note that after passage through the receiver this narrowest pulse will be just double the width of the original R.F. pulse from which it is derived, its width in its original form being 2/Bμsec.

24. When the pulses passing through the receiver are varying in width, the slopes of the leading and trailing edges produced by the band cutting will be the same for all pulses, and the absolute widening or narrowing of each pulse relative to its unmodulated width will therefore be unaltered. Thus, in spite of the degeneration of the shape of the pulses, the modulation will be carried through without change.

25. The above arguments show that the band-width of a pulse receiver is virtually decided once the narrowest pulse to be passed through has been chosen. It is quite usual to allow a considerable margin over and above this minimum band-width, and with such a margin the degeneration of the pulse edges will be reduced and the time taken to trace out the leading and trailing edges will be less than that for a receiver fulfilling only minimum requirements. The advantage of this will become apparent when the question of signal-to-noise ratio is later discussed.

Depth of modulation

26. The factor in a P.W.M. system which corresponds to carrier amplitude in an A.M. system is mean pulse width. The depth of modulation (quoted as a percentage) in A.M. is defined as (M/C) $\times$100%, where M is the maximum extent of the amplitude change from that of the unmodulated carrier of original amplitude C (see Fig. 10). The depth of modulation of a width-modulated pulse is similarly (d/a)$\times$100%, where a is the width of the unmodulated pulse and d is the extent of the time displacement of the edge suffering modulation. From conclusions reached in para. 23 it is clear that the depth of modulation employed in a pulse system of this type is limited by the band-width of the receiver once the mean pulse width has been decided upon. The band-width of the receiver is a function of the narrowest pulse to be passed,

and if the modulating process is allowed to produce pulses narrower than this limit, then they will suffer amplitude distortion in passing through the receiver, with consequent distortion of the intelligence they convey.

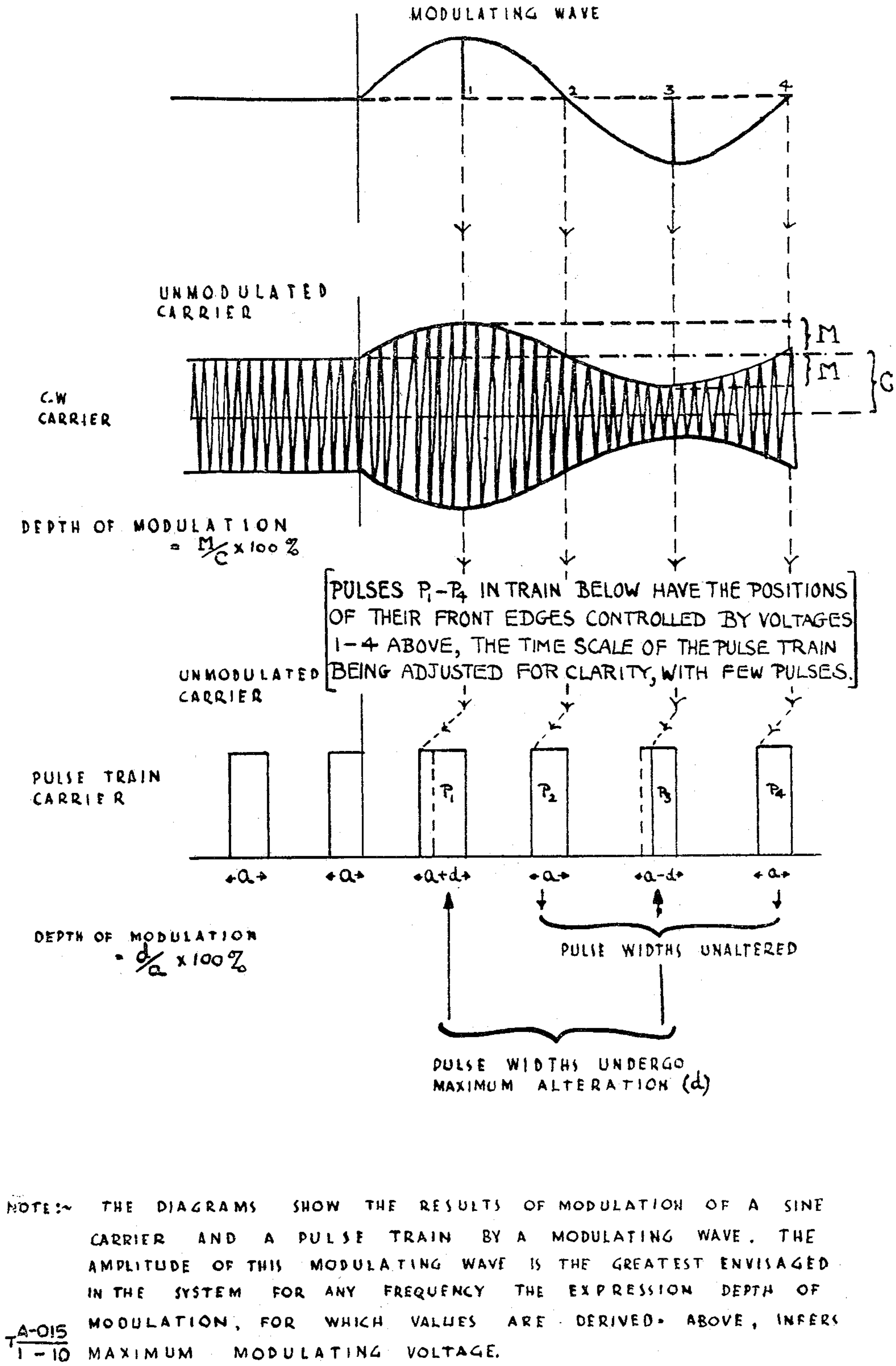


Fig. 10—Method of calculating depth of modulation

Signal-to-noise ratio in a P.W.M. system

27. In literature dealing with A.M. systems the expression "signal-to-noise ratio" is frequently used in rather a loose sense, and it is often not made clear in the context whether the ratio referred to is taken at the input to the R.F. stages, or at the output of the A.F. stages, or at what depth of modulation the measurement is made. Care will be taken in the following to specify the points at which the measurements are made and, where necessary, to specify the signal conditions at the time of observation. The ratio will always be taken to be that of the *voltages* concerned. Two typical transmission systems will be considered, both wireless links, one an A.M., the other a P.W.M. system. The following assumptions, physically realizable, are made in order to put the systems on the same initial footing for purposes of comparison :—

(*a*) The operating frequencies are the same.
(*b*) The aerial systems are similar.
(*c*) The receivers, apart from band-widths, are similar up to the first detector. The band-width of the A.M. receiver is $2f_m$ c/s, where f_m is the highest modulating frequency to be passed, and the band-width of the P.W.M. receiver is $10^6 \times B$c/s, where 2/B is the width in microseconds of the narrowest pulse to be passed.
(*d*) The transmitters have equal *mean* powers.
(*e*) The transmitters are modulated by the same audio intelligence frequency, in each case to a depth which will give the same audio output power at the first detector.
(*f*) The audio output circuits of both receivers at which the signal-to-noise measurements are taken do not change the ratios from those at the corresponding first detectors.

In such systems we may compare the signal-to-noise ratios at the audio outputs by the following reasoning. Let us consider the transmission of one complete cycle of the modulation frequency by the two systems. Since the incidence of noise voltages, apart from casual static, has a random distribution in both time and frequency, integrated noise power at any time is a function of the frequency selective properties of the circuits concerned, and is, in fact, directly proportional to the band-width. Therefore, in the case of the A.M. receiver, the signal energy and the noise energy (the latter being proportional to $2 f_m$), will be continuously distributed throughout the modulation cycle, and the signal-to-noise ratio at the audio output will therefore be a function of the ratio signal energy per modulation cycle/noise energy per modulation cycle. Similar considerations will produce the factor controlling the signal-to-noise ratio at the audio output of the P.W.M. system, although in this case the signal energy is not continuous, but "lumped" into pulses, which may be of short duration compared with the time interval between them. The output circuits of the P.W.M. system will, however, resolve this discontinuity into a continuity and the final audio output signal-to-noise ratio for the P.W.M. receiver will be arrived at in the same way as for the A.M. receiver, taking into account the fact that the band-width of the pulse receiver is $10^6 \times B$ c/s, and that the noise energy per modulation cycle will be proportional to this. For pulses of a few microseconds duration, because of the comparatively large receiver band-width required, this may give a total noise energy per modulation cycle considerably greater than that for the equivalent A.M. system, with the consequent production of a signal-to-noise ratio at the audio output of the P.W.M. receiver that is a good deal smaller than that of the A.M. receiver.

28. Thus, for example, let the A.M. band-width be 10kc/s and let the P.W.M. system employ a minimum pulse width of 10μsec., requiring a receiver band-width of 2×10^5c/s. Both systems are reasonably representative. The noise powers will be in the ratio (P.W.M. noise power)/(A.M. noise power) $= (2 \times 10^5)/10^4 = 20/1$, and the noise voltages in the output stages will therefore be in the ratio $(20/1)^{\frac{1}{2}} = 4.45/1$. Thus, under the conditions of comparison stated at the outset, the signal-to-noise ratio at the output of the P.W.M. receiver will be only 1/4.45 of that at the output of the A.M. receiver.

29. The above systems were compared under conditions of equal mean power. In the case of a P.W.M. system this would mean, for comparatively narrow pulses, a peak power many times greater than the mean power, as shown graphically in Fig. 11. On the other hand, it is a feature of P.W.M. systems that the peak power required to operate them is to a very large extent independent of the pulse width, and that, in fact, the mean power of the transmitter may be considerably less than that used in an equivalent A.M. system. With such a reduced value of signal power, since the signal energy delivered to the receiver per modulation cycle may now be only a fraction of that contemplated in the above comparison, while the noise energy per modulation cycle will be unaltered, *it might appear that the signal-to-noise ratio at the audio output of such a system would compare so unfavourably with that of an equivalent A.M. system as to militate against its successful use under similar field conditions, except with comparatively wide pulses.* It is, nevertheless, the ability to use P.W.M. with comparatively narrow pulses that renders it of especial value in multi-channel systems (paras. 60 to 65), and it will now be shown how this difficulty can be overcome, and how the signal-to-noise ratio for a P.W.M. system can be made equal to, and even better than, that of an equivalent A.M. system.

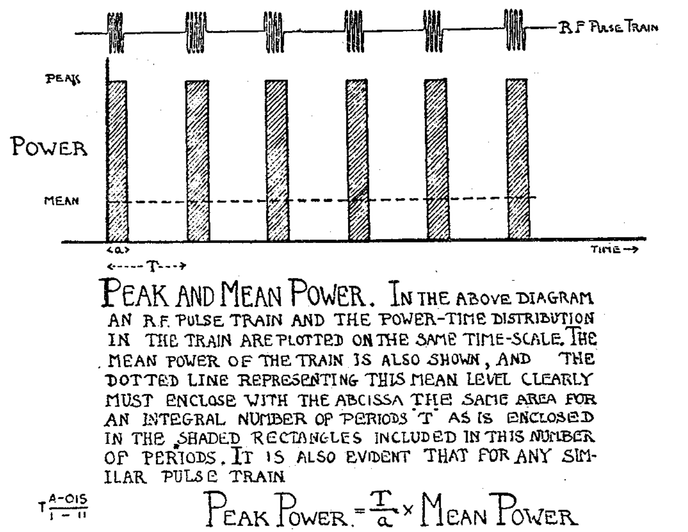


Fig. 11—Ratio of peak and mean powers

30. In order to understand how this improvement can be effected, it is necessary to understand the action of a device, common to all pulse systems except those using amplitude modulation, and known as the "slicer" or "noise slicer." It will be best to study the action of this device from the aspect of its treatment of a theoretical square pulse, and then to discuss its practical limitations. Let us suppose, then, that, at the transmitting terminal of a P.W.M. wireless link, a D.C. width-modulated pulse train has been used to modulate an R.F. transmitter and that resultant square R.F. pulses have been produced. It will be assumed for the moment that the receiver is capable of passing these pulses, in their R.F. or corresponding I.F. form, with so little degeneration that after amplification and detection they appear as D.C. pulses at the output of the receiver first detector in their original geometrical form with only the addition of noise voltages acquired in passage through the tuned stages of the receiver. Let these noise voltages be those due to thermal agitation and valve effects and not to casual static. Now, to an extent dependent upon the amplitude of the signal at the receiver input, and upon the receiver band-width, the detected pulse train, when presented upon a C.R.O., will be seen to be associated with these noise voltages, which will have the usual random distribution in both time and frequency. The physical form of the trace will be as shown in Fig. 12 (*a*). If the detected pulses are still perfectly square in form, the rise from zero to maximum will take place instantaneously and therefore no noise voltages will appear upon the sides of the pulses.

31. If it is now possible to select for amplification and subsequent detection of the pulse-width modulation only that portion of the pulses which lies between two voltage levels indicated by the horizontal lines of Fig. 12 (*a*),

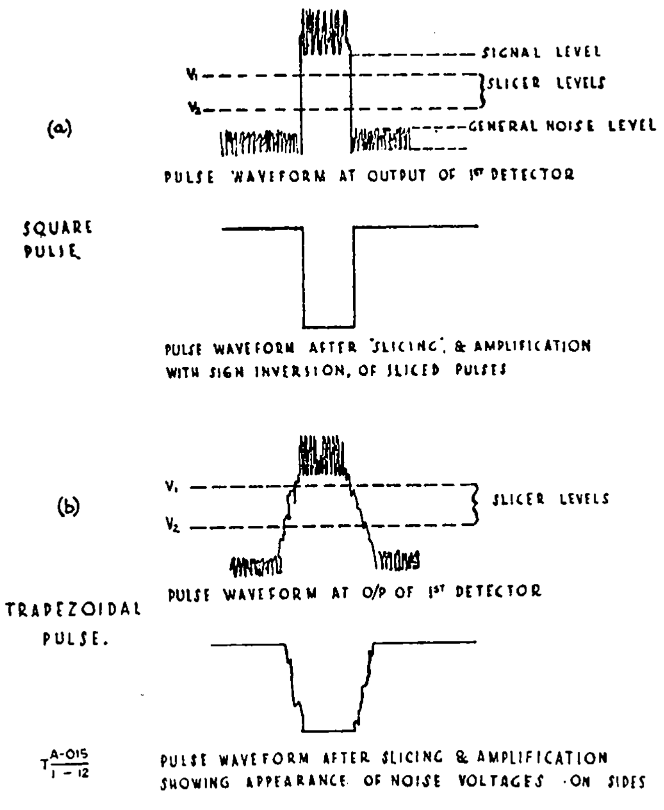


Fig. 12—Operation of slicer

the noise voltages will be rejected and will not appear in the subsequent wave form. This is made possible by applying the wave form to a valve having a short grid base and linear characteristic in such a way that the positive end of each pulse is squared off by grid current or saturation, while the negative end carries the valve beyond cut-off. Thus, for a theoretical squ[illegible] pulse, the result of this slicing would be the removal of all noise content. Circuit details of a typical noise slicer are given in para. 56.

32. Now in practical cases, as was shown in para. 23, the detected pulses are never square, but trapezoidal, in form, this arising from the necessity of limiting the band-width of the receiver, with consequent degeneration of the pulse edges. Such pulses can clearly have noise voltages superimposed upon their sloping sides, as in Fig. 12 (*b*), and this will result, after the slicing process, in random fluctuations in pulse width over and above the deliberate variations due to modulation, this producing the usual effects of noise in the audio output circu[illegible] Thus, while pulse transmission systems do lend themselves to a reduction in the value of the noise present in the final signal, it must not be supposed that complete suppression of noise can be achieved in any practical system. The theoretical possible improvement in the noise present in the output for a given level of signal over that present in an A.M. receiver operating under similar conditions is limited by such considerations as the above, and these are discussed from a mathematical point of view in paras. 85 to 93.

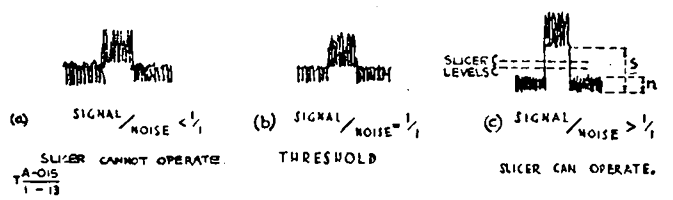


Fig. 13—Ratio ne[illegible]ssary for operation of slicer

33. Before proceeding to give a quantitative comparis[illegible] between the audio output signal-to-noise ratios for equivalent A.M. and P.W.M. receivers operating under similar conditions, it is necessary to define "threshold" conditions at the R.F. stages of such receivers. In A.M. systems threshold is defined as the signal level for which, at the audio output stages, the signal-to-noise ratio is 1/1. In a P.W.M. system threshold conditions can be defined in a similar way if the output of the first detector is made the basis of the definition. It should be noted that the expression signal-to-noise ratio in this context means the ratio s/n of Fig. 13 and has no bearing on the audio signal voltage which might be developed from such a pulse. The reason for this definition of threshold conditions will become clear if reference is made to Fig. 13. Here the voltage wave forms for three typical cases are shown after the first detection process. In Fig. 13 (*a*) the signal-to-noise ratio is less than 1/1, in Fig. 13 (*b*) the ratio is exactly 1/1, while in Fig. 13 (*c*) it is greater than 1/1. Now any one of the pulse modulation methods, except that of amplitude modulation, depends for success upon reasonably precise definition of the edges of the pulses. The condition of Fig. 13 (*b*) is such that a very slight [illegible]crease in the signal level will raise the sides of the pulse clear of the general noise level and allow the noise slicer

to operate satisfactorily. In Fig. 13 (*c*) the sides of the pulse are clearly defined and adequate conditions exist for good reception. Thus, in general, *threshold conditions may be said to exist in a P.W.M. system when the signal-to-noise ratio at the output of the first detector is* 1/1. For good reception something better than this is required, and a signal-to-noise ratio of 2/1 at this point is frequently quoted as a criterion of good operation. It should be added that although the noise slicer will commence to operate at a signal-to-noise ratio just greater than 1/1, increase in pulse amplitude above the value to produce this limiting ratio will always produce a corresponding increase in the final signal-to-noise ratio at the audio output, for reasons which may be seen in paras. 86 to 90.

34. Once the signal amplitude in a P.W.M. system is sufficient to allow the noise slicer to operate correctly, then, from an operational point of view, the signal-to-noise ratio at the output of the first detector has less importance than the signal-to-noise ratio at the audio output. Because, as will be shown in paras. 86 to 94, it is primarily the receiver band-width which determines the final audio signal-to-noise ratio, the factor which compares the signal-to-noise ratios at the audio outputs of P.W.M. and A.M. systems operating under similar conditions and with the same mean power is known as the "wide-band gain" of the pulse receiver. This factor is arrived at by stipulating that signal conditions at the P.W.M. receiver must allow the noise slicer to operate correctly, and by then laying down all the essential conditions under which the comparison is to be made, such as depth of modulation, highest modulating frequency and band-width for the A.M. receiver, and depth of modulation, mean pulse width, band-width, pulse recurrence frequency and highest modulating frequency for the pulse receiver. The systems are arranged to be working under conditions which give optimum values for the signal-to-noise ratios at the audio outputs, and with the same mean power. The result obtained is that :—

Wide-band gain =

$$\frac{\text{S/N ratio at audio o/p of P.W.M. system}}{\text{S/N ratio at audio o/p of A.M. system}} = \frac{(3aB)^{\frac{1}{2}}}{4} \cdot \frac{d}{a}$$

where B is receiver band-width in Mc/s, a is the mean pulse width in microseconds and d is the deviation required to give the narrowest pulse allowable.

35. Thus, for example, consider a system employing a pulse of minimum width, even narrower than that considered earlier in the text and requiring therefore a correspondingly greater band-width. Let the unmodulated pulse be 20μsec. wide, diminishing to a least width of $\frac{1}{2}\mu$sec. for full modulation. This will require a receiver band-width of 4Mc/s, and these figures, when substituted in the above expression, give a wide-band gain of 3.77, showing that, in fact, this particular P.W.M. system, with slicer, is very much better than an equivalent A.M. system. It is also interesting to compare the powers required by the two equivalent systems to give the same final signal-to-noise ratios. We may clearly reduce the power of the P.W.M. system in the ratio $(1/3.77)^2$ since the signal-to-noise ratios are measured in terms of output voltage, and this produces the result that the P.W.M. system requires a mean power of only about 1/14th. of that of the equivalent A.M. system to give similar reception.

36. If the narrowest pulse to be passed through the receiver is of width (a–d) μsec., and the band-width of the receiver is always adjusted to pass this narrowest pulse, then the expression for wide-band gain may be written :—

$$\text{Wide-band gain} = \frac{\sqrt{6}}{4} \times \sqrt{\frac{d^2}{a(a-d)}}, \text{ since now } B = \frac{2}{(a-d)}$$

and from this it will be seen that for a chosen mean width of pulse it is advantageous to make d as large as possible, i.e., to use the greatest possible depth of modulation. Then the wide-band gain becomes limited only by the practical difficulties associated with generating and transmitting very narrow pulses, and of designing a receiver to pass them.

37. Suppose now that (a–d), the narrowest pulse, has been fixed by such practical and economic considerations, so that, whatever the mean width of the unmodulated pulse, the quantity (a–d) may not now be altered. The expression for wide-band gain may now be rewritten :—

$$\text{Wide-band gain} = \frac{\sqrt{6}}{4} \times \sqrt{\frac{(a-k)^2}{a.k}} \text{ where } k = (a-d). \text{ It}$$

now appears that the greater a becomes, the greater the resulting wide-band gain. This fact will be discussed in the following paragraph.

Choice of pulse width in a P.W.M. system

38. From the results of the above discussion it is now possible to outline the factors deciding the mean pulse width of a P.W.M. system. For a given *mean power* and pulse recurrence frequency it appears that the wider the unmodulated pulses, the greater will be the wide-band gain. On the other hand it must be remembered that the advantages of a P.W.M. system in this respect are realized only if the slicer operates correctly, i.e., if threshold is passed, and thus where considerations of transmitter design and distance are likely to limit the signal level at the receiver rather severely, it will be better to use narrower pulses, and so secure the advantage of greater pulse height. Again, where the P.W.M. system has to fulfil multi-channel requirements, as will appear in paras. 62 and 63, the time interval between successive pulses becomes of as great importance as the duration of the pulses themselves, because it is required for the time-interlacing of the pulses of other channels of communication. In such circumstances it is necessary to sacrifice the advantages of a large possible wide-band gain and to use comparatively narrow pulses in order to secure the multi-channel facility. In spite of this sacrifice practice shows that it is quite possible to have a large number of channels operating over a single wireless link, with a mean power per channel not greater than that of an equivalent single-channel A.M. system, and with a wide-band gain appreciably greater than unity.

Choice of pulse recurrence frequency in a P.W.M. system

39. The mean pulse width having been decided, it remains to choose the pulse recurrence frequency. The choice of this factor mainly depends upon the number of channels to be employed, the power available at the transmitter and the highest modulation frequency to be transmitted. For a given mean transmitter power, and for a single-channel system, the smaller the number of

pulses to be transmitted per second, the greater may be their relative amplitude, with corresponding increase in range and/or wide-band gain, optical path considerations frequently setting the upper limit of the range where U.H.F. transmitters are used. On the other hand, since the pulse recurrence frequency occurs as a component in the audio output wave form, unless a second detector and filter are employed, considerations of simplicity in receiver design may make it necessary to keep the pulse recurrence frequency above the audio spectrum, so avoiding a continuous whistle in the output. Thus, in such systems the pulse recurrence frequency might be set at, say, 20kc/s to give a fair compromise.

40. Where as large as possible a number of channels are required over the P.W.M. link and the time-interlacing method is to be employed it is necessary to space out the pulses of any one channel to the greatest possible extent on the time axis to " make room " for the interlaced pulses of other channels (paras. 62, 63). Thus it becomes necessary to work with the lowest possible pulse recurrence frequency, and the lower limit of this frequency is set by the highest intelligence frequency to be transmitted. Let us suppose that 3kc/s is the highest modulating frequency, and that 9kc/s has been chosen as the recurrence frequency. Then for the highest modulating frequency one complete cycle of intelligence has to be conveyed on three pulses. The net effect of this is shown in Fig. 14 (*a*), and it will be seen that modulation is quite possible, whatever the phase relationship between the intelligence wave form and the pulse carrier. The detection of the pulse-width modulation will ultimately convert the width variations of the pulses into corresponding voltage variations, and appropriate voltages are shown plotted on the same scale. The fundamental frequency of these voltage variations is 3kc/s, and this frequency can be filtered off from the associated frequencies by a suitable low-pass filter. Now suppose that the pulse recurrence frequency is lowered to 6kc/s. Fig. 14 (*b*) shows that it is now possible for the phase relationship between the modulating wave and the pulse train to be such that the following cases may arise :—

(*a*) Maximum modulation is obtained.
(*b*) No modulation results.
(*c*) Some intermediate value of modulation is obtained.

Finally, consider the effect of lowering the pulse recurrence frequency to 3kc/s. Fig. 14 (*c*) now shows that no modulation is possible in any circumstances. It is thus shown, in this tentative treatment, that *the pulse recurrence frequency should be at least three times the value of the highest modulating frequency for successful modulation.* Since 3kc/s is the normal upper limit set to speech frequencies for telephone working, for a simple speech transmission system the lowest limit of recurrence frequency will be 9kc/s.

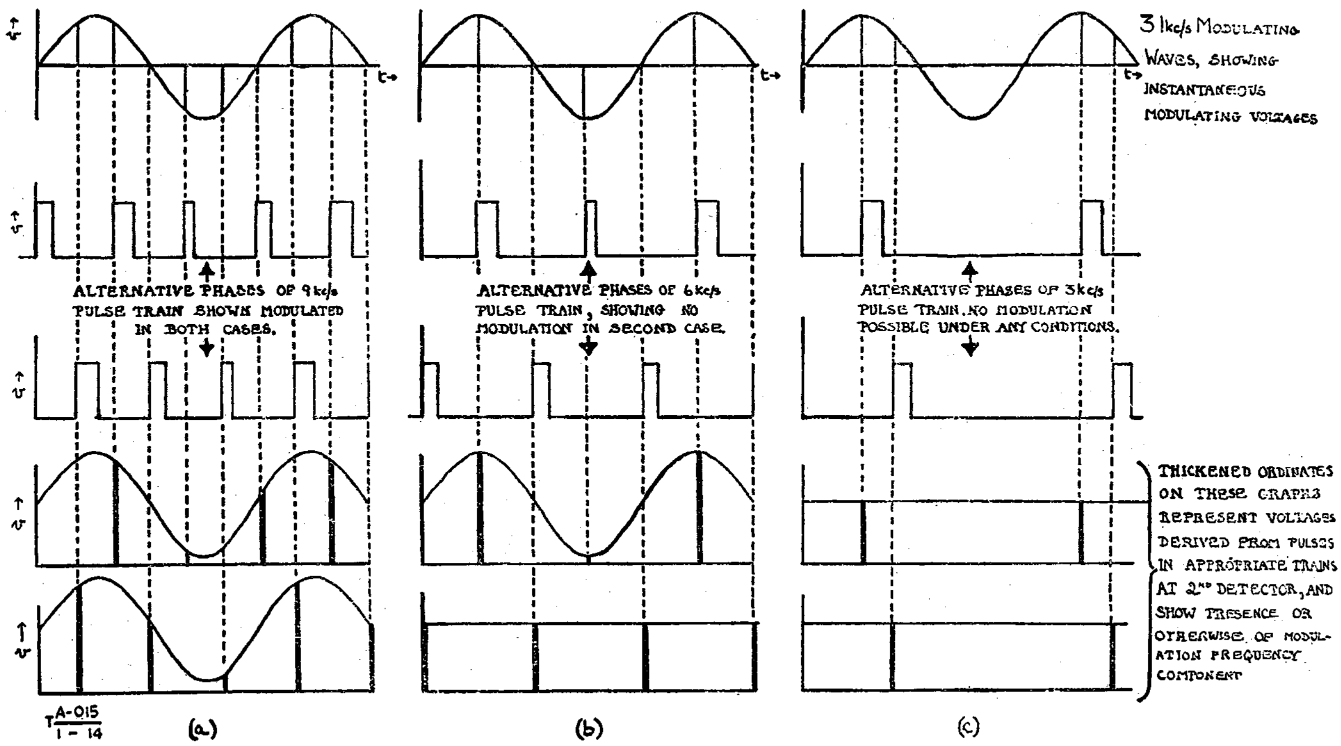


Fig. 14—Diagram showing results when carrier frequency approaches modulation frequency

MISCELLANEOUS SUBJECTS

Transmitter valves

41. The employment of narrow pulses, with their comparatively high component frequency content, necessitates the use of high radio frequencies (and intermediate frequencies) so that the problems concerned with passing these pulses through the receiver up to the first detector may not be insoluble. As a direct result of this it is necessary to transmit on V.H. or U.H. frequencies. Now it is a general feature of transmitters operating at these high frequencies that stability of frequency is difficult, if not impossible, to obtain, especially when wavelengths of less than one metre are envisaged, and crystal control by frequency multiplication methods becomes impossible owing to transit time and other effects. On the other hand, considerations of privacy, efficiency, and compactness have changed the general requirements from the metric to the centimetric band. The existence of well-tried types of oscillator valves, such as magnetrons, kylstrons, etc., capable of producing centimetre oscillations, but with special peculiarities, makes the use of some form of modulation with a constant-voltage amplitude characteristic a necessity. Thus a magnetron requires a reasonably constant optimum voltage to oscillate successfully, or even to oscillate at all. Small changes in H.T. voltage produce quite apparent changes in frequency, and no successful method of amplitude modulation is yet possible. The solution at the present item must therefore be modulation by means of constant-amplitude voltage pulses. In spite of this method of modulation frequency drift of the oscillator is a major problem in a radio link employing centimetre wavelengths, since slight changes in circuit parameters at the oscillator and modulator are unavoidable, and are always accompanied by corresponding changes in carrier frequency. To take account of such casual drift, two features are usual at the pulse receiver, namely, the provision of a band-width rather wider than that strictly necessary from considerations of the central transmitter frequency and the width of the narrowest pulse, and the inclusion of some system of automatic frequency control of the local oscillator at the receiver, so that it " pulls itself into tune " as the transmitter (or local oscillator) frequency drifts.

Aerial systems

42. The necessity of using metric and centimetric wavelengths with the above consequent disadvantages does, however, bring with it the complementary advantages of simplicity in aerial design and ease with which beaming may be carried out by the use of stack arrays or mirrors. Such aerial systems will usually give a quite large aerial gain and so render the system not only more directional but also more efficient. It is only fair to say, however, that, with decreasing wavelength, the possible range of the system becomes increasingly limited, tending towards that of a purely optical path in the centimetric band.

Privacy

43. The use of a pulse-modulated transmission system is not, of itself, a guarantee of privacy, since almost any type of radio receiver capable of receiving the radio frequency will produce, with varying degrees of efficiency, a voltage variation at the detector circuit which contains the intelligence frequencies as components. Other factors, however, tend to make pulse systems more private than normal A.M. systems. The use of centimetre wavelengths, with its consequent possibility of beaming, and restriction of path to optical range, makes casual interception unlikely. Where the utmost privacy is desired, and the radio frequencies do not permit of beaming, it is often possible to superimpose a spurious amplitude modulation on the width-modulated pulse train, and so render the intelligence unintelligible except in a receiver employing a slicer. Finally, in multi-channel systems, the intelligence of one channel will mask that of another channel where casual reception without gating is employed. In such systems, when only one channel is in use, it is possible to modulate the channels not in use by spurious signals to improve the degree of privacy on the channel in use.

Jamming

44. Discussion of the full possibilities of jamming are beyond the scope of this treatment. It may be pointed out that the factors which tend towards privacy, i.e., beaming and limitation of path to optical range, tend also to limit the possibility of jamming. Where gating is employed a further safeguard against all but C.W. jamming is inherent in the system, and the power necessary for such interference completely to saturate the receiver and render the noise slicer inoperative would be very great.

SUMMARY OF ADVANTAGES AND DISADVANTAGES OF P.W.M. SYSTEMS

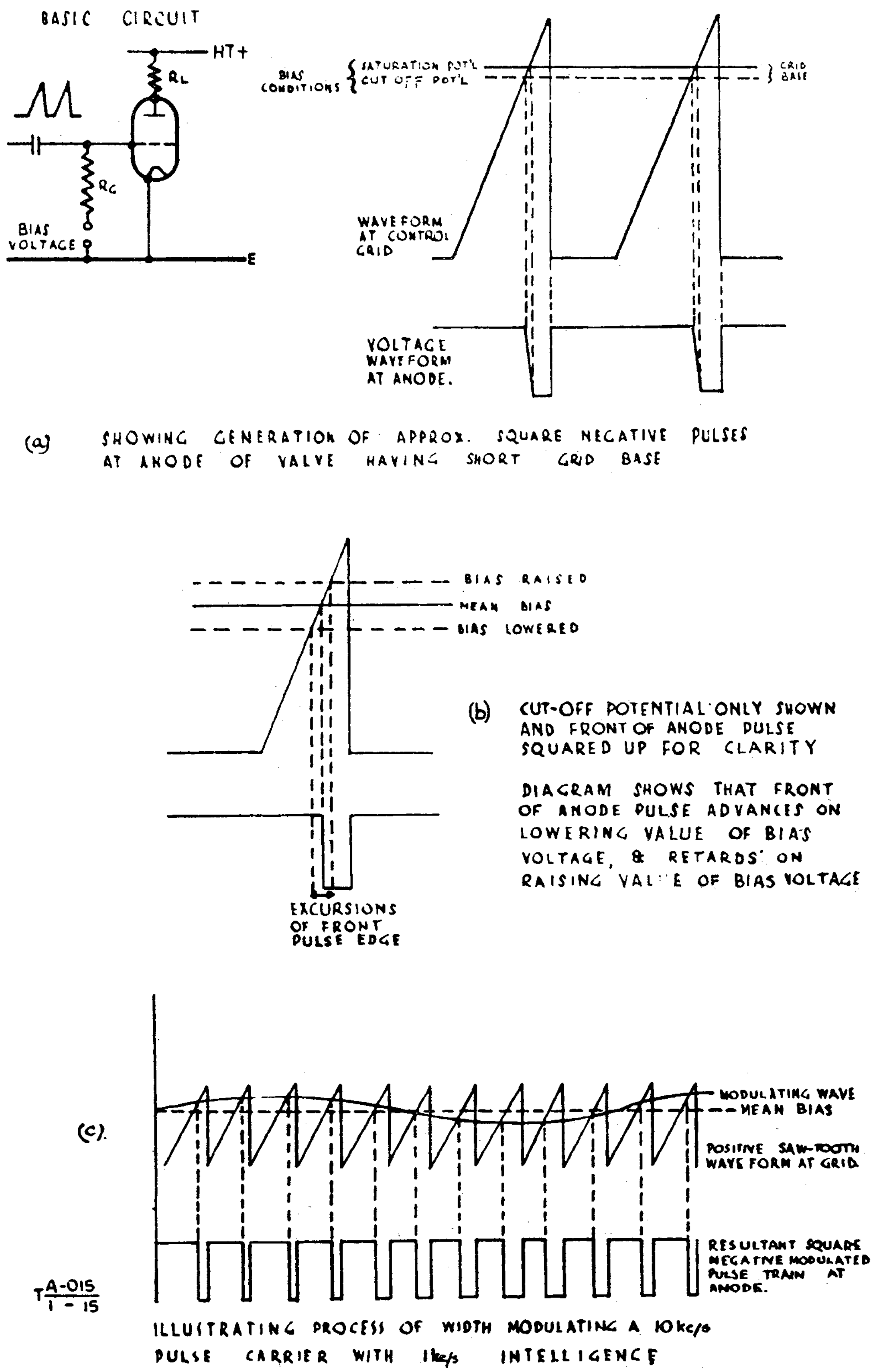
45. The *advantages* are as follows :—

(*a*) Possibility of large wide-band gain.
(*b*) Possibility of using types of valves (for production of short waves) which do not permit of A.M.
(*c*) More efficient operation of valves because of ability to fix constant optimum operating voltage.
(*d*) Multi-channel possibilities.
(*e*) Clearly defined threshold conditions, with consequent possibility of economical use of valve power.
(*f*) Privacy (limited improvement possible).

46. The *disadvantages* are as follows :—

(*a*) Relatively wide band required at the receiver.
(*b*) Increase in weight and complexity of equipment owing to necessity of using two processes of modulation, and possibly two of detection, and of having a slicer before the final output.

In connection with these disadvantages it is worth while noting that the question of receiver band-width is not an acute one, since radar developments have made the design and construction of wide-band receivers one of comparative simplicity. With regard to the possible increase in weight and complexity of equipment, it is not found in practice that this does in fact result, since the ability to operate in the centimetre band and with low-power transmitters results in a decrease in weight and in the complexity of the high-frequency circuit details since R.F. stages are now no longer possible.



T A-015 / 1 - 15

Fig. 15—Production of width-modulated pulses

CIRCUITS FOR PRODUCING AND SHAPING WIDTH-MODULATED PULSES

47. The principles underlying the most general method of producing width-modulated pulses may be summarized as follows. A linear sawtooth voltage wave form, recurrent at the eventual pulse recurrence frequency, is fed to the control grid of a valve which has a linear anode-current, grid-voltage characteristic, and which operates with a comparatively short grid base. In the absence of a modulating signal the valve is so biased that the instant of conduction occurs at a fixed point in the progress of the sawtooth wave form. The shortness of the grid base, which causes the valve effectively to saturate soon after the applied wave has carried the grid through cut-off potential, results in the production, in a resistive anode load, of an approximately rectangular negative-going wave form of voltage. Fig. 15 (*a*) shows this process for a typical applied sawtooth wave.

48. If now the standing bias on the valve be raised or lowered, a variation in the width of the anode pulse will result, and this variation will from geometrical considerations, be directly proportional to the variations in the bias voltage. The point will become clear by reference to Fig. 15 (*b*) where to avoid confusion the grid base of the valve has been reduced to zero.

49. If then, during the application of a train of pulses of sawtooth form at the grid, the standing bias on the valve be raised and lowered in accordance with the configuration of a voltage wave form of intelligence, the widths of the anode pulses will vary linearly in accordance with the values of the instantaneous modulating voltages, and these voltages are selected from the wave form in the manner outlined in para. 15. Reference to this paragraph, and a study of the mechanism of anode pulse production in the above circuit, will show that the straight line graph of Fig. 7 (*b*) is identified with the front edge of the sawtooth form of the above circuit. Paras. 77 to 85 show that, although the instantaneous modulating voltages are not spaced out regularly in the modulating wave by this method of modulation, the width modulation is, nevertheless, linear. Fig. 15 (*c*) shows a 1kc/s intelligence frequency modulating a 10kc/s pulse carrier, and the train of width-modulated pulses so produced.

50. A basic circuit for producing modulation in this way is shown in Fig. 16 (*a*). It consists of a valve whose cathode potential is fixed, in the absence of a modulating signal, at the required potential by means of a potentiometer between the H.T. supply rails. The sawtooth wave form is fed direct to the control grid and the intelligence voltage to the cathode. Raising the cathode potential will result in narrowing, and lowering the potential in widening, the pulse at the anode of the valve. The intelligence thus width-modulates the pulse train.

51. Another form of basic circuit is shown in Fig. 16 (*b*). Here the sawtooth form is coupled into the control grid and the potential of this grid is also varied by the modulating voltage which is fed to the free end of the grid leak resistor. The initial width of the unmodulated anode pulse may in these circumstances be adjusted by the selection of the time-constant of the coupling components, which allow slide-back action to produce a standing bias at the grid of such value that the valve commences to conduct at the correct point in the progress of the sawtooth. This type of circuit has been used in equipment where a wide mean pulse width is allowable and where, consequently small long-term variations in the mean width of the anode pulses due to changes in circuit parameters have little consequence. Where, however, a narrow pulse is employed and equal variations would now produce very much more serious effects, a modified form of the circuit of Fig. 16 (*a*) has been used with success. This is shown in detail in Fig. 16 (*d*). It is not advisable to fix the cathode bias potential by means of a potentiometer, since then long-term variations in the amplitude of the sawtooth wave form would produce serious variations in the mean width of the anode pulses. Instead, the cathode bias is made automatic and is derived by the use of a cathode resistor and by-pass condenser. Then, if the valve cathode resistor has a comparatively high value and the anode load is small, we may neglect for the moment the action of the by-pass condenser in considering the action of the circuit. The progress of the positive-going edge of the sawtooth wave at the control grid will produce a similar wave form of rather smaller amplitude at the cathode, the valve now behaving as a cathode follower. Thus, at the culmination of the applied sweep the cathode potential will be only a few volts below that of the grid, the exact difference in voltage being a function of the amplitude of the grid sawtooth wave. Now the effect of inserting a decoupling condenser in the cathode circuit will be such that, although the development of the positive edge of the sawtooth form at the cathode will not be impeded for the first pulse, the high final potential will be maintained due to the inabilty of the decoupling condenser to discharge through the cathode resistor before the next sawtooth wave appears at the control grid; and so the trailing edge of each sawtooth wave will cut the valve off, and it will not be opened up again until the positive edge again carries the grid up to the cut-off potential. The pulsing action of the wave form at the grid will maintain a high and steady bias at the cathode and the valve will thus conduct only on the peaks of the applied sawtooth, the action being illustrated in Fig. 16 (*c*). The width of the anode pulses in the unmodulated condition will thus be determined primarily by the difference between the peak voltage at the grid and the standing cathode bias, and this difference will itself be a function of the amplitude of the applied sawtooth wave form. The two main results of this action are:—

(*a*) Small changes in the amplitude of the applied sawtooth wave form do not produce very much change in mean pulse width, since the cathode potential always adjusts itself automatically to the amplitude of the grid wave form.

(*b*) The mean pulse width is continuously variable by means of adjustment of the amplitude of the applied sawtooth wave form, and thus a mean pulse width may be accurately set in before modulation is applied.

52. If now the cathode can be made to vary in potential according to the modulating voltage, then width modulation of the anode pulses will result. But the cathode is decoupled for the pulse recurrence frequency, and if the highest modulating frequency is near to this, it, too, will be decoupled. Thus where these two frequencies fall close to each other, the time-constant of the decoupling circuit has to be chosen with great care, and at the best it can be only a compromise, allowing a high bias to be maintained and at the same time allowing as full a development of the modulating wave form as possible. The values given in

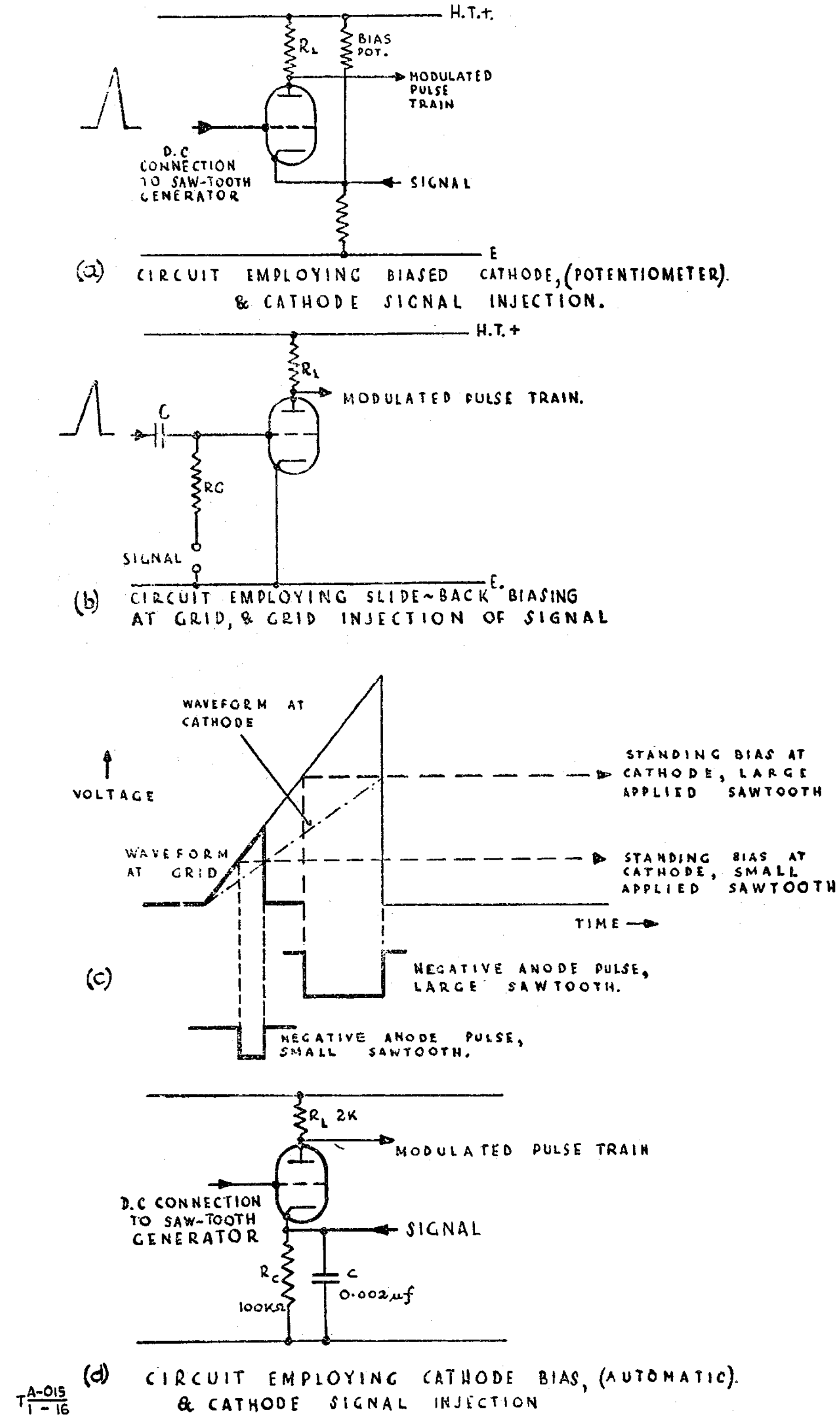


Fig. 16—Basic circuits for producing pulses

Fig. 16 (*d*) are those used in the W.S.10 equipment, where the pulse recurrence frequency is 9kc/s, and the highest modulating frequendy is 3kc/s, and where, therefore, the time-constant of the cathode circuit can vary only between very narrow limits. In spite of careful choice there is bound to be some degeneration of the higher frequencies of the intelligence and frequency non-linearity is introduced.

Shaping the modulated pulse train

53. The defects of the pulses generated and modulated by the last method will now be considered together with the remedies used to mitigate these defects. The defects are :—

(*a*) The leading edges of the anode pulses are sloping owing to the definite time interval needed for the sawtooth wave form at the control grid to carry the valve from cut-off to full conduction, this defect being shown in Fig. 15 (*a*).

(*b*) The tops of the pulses will be rounded due to the failure of the valve to saturate instantaneously.

(*c*) The pulses will be amplitude-modulated for the following reason. Variations in cathode potential will vary the drop in potential at the anode when the valve conducts by varying the *effective* H.T. voltage on the stage. Thus a high cathode potential will allow a smaller fall in anode voltage than a low cathode potential.

These defects of the embryo pulses may be readily remedied by application to standard pulse-shaping circuits. The processes involved, known as " clipping " and " slicing " can now be described.

Pulse clipping

54. This process may be applied to both positive and negative voltage pulses. That for negative pulses will be described first since it is of most immediate interest Reference is made to Fig. 17. The train of pulses is fed to the grid of a valve, the grid being returned to a point of high potential through a leak of high resistance. A very slight flow of grid current results in nearly all the high applied potential being balanced out by the drop across the leak, with the result that the grid is held just slightly positive with respect to cathode potential. The valve is chosen to have a short grid base and linear grid-voltage, anode-current characteristic down to cut-off, so that only a small part of the negative voltage excursions of the pulses is effective in causing a current change through the valve, as shown in Fig. 17 (*b*). This has the result that the duration of the front edge of the pulses reproduced in the resistive anode load is considerably shortened, since any part of the pulses executed beyond cut-off is effectively discarded. For the same reason all pulses, whatever their amplitude at the grid, provided that they are large enough to carry the valve beyond cut-off, will produce constant-amplitude anode pulses, and the tops of the pulses so produced will be flat. The voltage excursion of these positive-going anode pulses will depend upon the gain of the stage, and may be made considerably larger than the grid base of the valve. Thus at one operation the amplitude modulation of the pulse train is removed, the pulses are " cleaned up," and the sides steepened. Fig. 17 (*c*) shows the results of the process, which may be repeated, after sign inversion of the positive train, to any degree required.

55. Positive pulse trains requiring clipping may be subjected to an equivalent process. The necessary circuit is shown in Fig. 18. The positive pulse train is applied to a valve having zero cathode bias, but with a sufficiently large time-constant at the grid circuit to ensure that the valve remains biased off (due to grid current flow) until the positive peaks of the applied pulses cause it to conduct. Thus only the peaks of the positive pulses are reproduced (as negative pulses) in the anode load and the reproduced pulses will have constant amplitude controlled by the gain of the stage. If the original pulses have rounded tops, this defect may be removed by the inclusion of a squaring resistor in the grid circuit. It will be seen that the action of this circuit is not quite so suitable for removing amplitude modulation as the process for negative pulses, since a small pulse following rapidly after a large pulse may not overcome the whole of the residual slide-back bias, thus still producing a correspondingly small negative pulse at the anode. This effect will always be reduced, but never completely overcome, by the inclusion of the squaring resistor mentioned above, which tends to limit the slide-back action.

Pulse slicing

56. Reference has already been made, in paras. 30 to 32, to the slicing process whereby effectively a slice of the given pulse is taken at about the centre of the voltage excursion for amplification. This process is particularly important as a means of ridding detected pulses of their noise content, so it will be described in detail for such pulses, although the circuits given may be used for any positive or negative pulse train. A typical circuit is shown in Fig. 19, together with diagrams of characteristic pulses before and after passing through the slicer. The valve is again a straightforward amplifier with linear characteristics and operating with a comparatively short grid base. The working-point voltage at the grid is adjustable by means of a potentiometer, so that the applied pulse wave form may be set, with respect to the grid base, in such a way that the negative part of the wave form is traced beyond cut-off, while the positive part is squared off by the action of the large-valued grid squaring resistor. In operation the bias at the grid is continuously adjusted until the output wave form shows that the amplitude modulation at the top and the base of the pulse train has been eliminated. It is evident that the extent of the clear sides of the pulses must equal the grid base of the valve before the slider can remove all the noise voltages. In addition to removing the noise voltages, this process also steepens up the edges of the pulses.

Circuits for detection of width modulation

57. The first detector performs the function of extracting the envelope of the R.F. or I.F. pulse train, after which the major part of the noise content is removed, and the pulses cleaned up by the slicing process. A wave form which, apart from slight random fluctuations in width due to the incidence of noise voltages upon the edges, conforms in every respect to the original D.C. wave form at the sender, now results. Such a wave form, containing, as it does, a component frequency equal to that of the modulating wave, may itself actuate the appropriate receiving device, and so reproduce the intelligence without further treatment. However, where the pulse recurrence frequency lies within the audio spectrum, it is usual to employ a second detector which converts the width

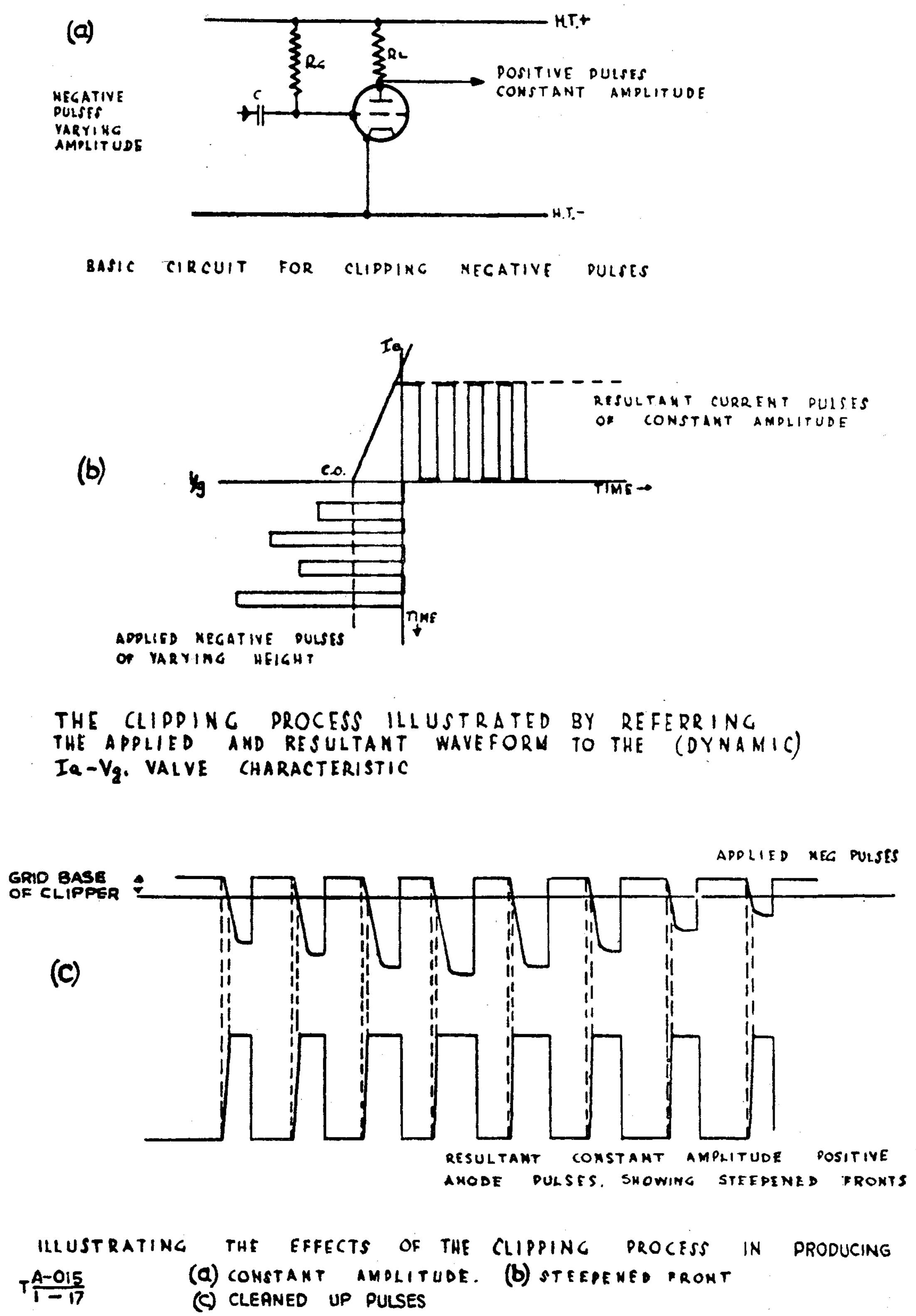
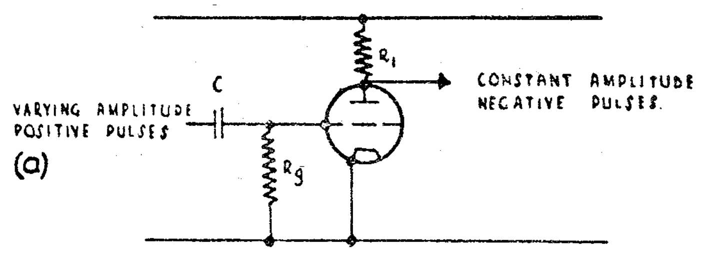


Fig. 17—Application of clipping circuit to negative pulses



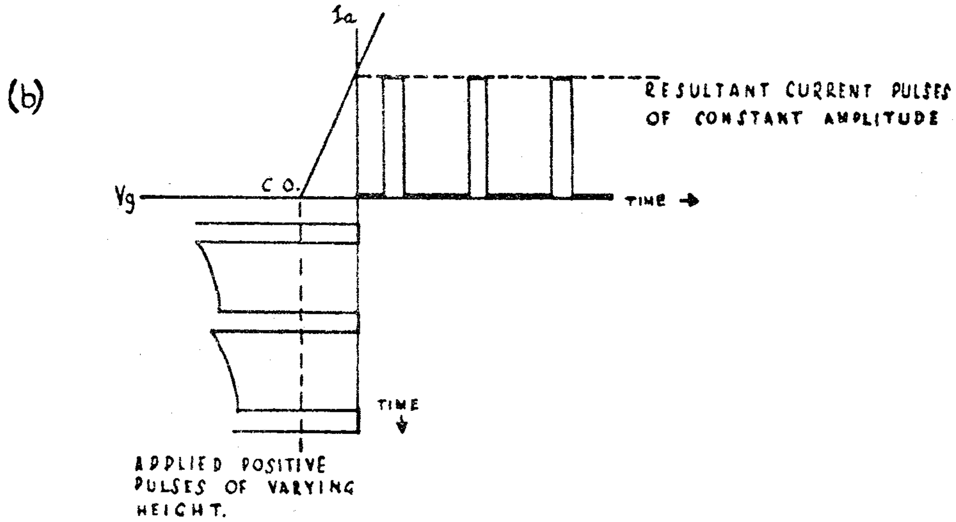


Fig. 18—Application of clipping circuit to positive pulses

variations of the pulses into corresponding voltage variations and at the same time reduces the relative amplitudes of the recurrence frequency component and its harmonics. The actual process involved is very simple indeed. It consists of applying the wave form to an integrating circuit (Fig. 20 (*a*)) of suitable time-constant so that the charging of the condenser is not completed during the passage of any single pulse, and so that, therefore, the voltage rise across the condenser is dependent upon the duration of the pulse (Fig. 20 (*b*)). The results of this integration process, for a pulse train modulated to a depth of 50% by an intelligence frequency one tenth of the pulse recurrence frequency, are shown in Fig. 20 (*c*). Here the voltage excursion of the condenser has been limited by careful choice of time-constant to the initial and most linear portion of the charging curve, and as a direct result of this the envelope of the output wave form will follow the form of the original modulating wave, since, as purely geometrical considerations will show, the height of each integrated pulse will be directly proportional to the width of the rectangular pulse generating it. This is ...ear pulse-width detection, and it is now necessary to remove the ripple, i.e., the pulse recurrence frequency and its harmonics. This may be done by passing the train of integrated pulses to a low-pass filter whose cut-off frequency is between the highest modulating frequency to be transmitted and the pulse recurrence frequency. The output of this filter will consisit of the intelligence only. The whole process is clearly analogous to normal detection of an I.F. wave form by a linear detector, followed by an I.F. filter, the detector serving to extract a wave form following that of the original modulating wave, and the filter to remove, in this case, the I.F. ripple.

58. It has been assumed in the foregoing description that the time intervals between successive pulses are sufficiently long to allow of full discharge of the integrating condenser between each pulse and the next in the train. Where this would not be possible with the simple circuit illustrated in Fig. 20 (*a*) the circuit of Fig. 20 (*d*) will ensure rapid and complete discharge of the condenser. A positive-going wave form applied to the circuit will cause the condenser to charge normally through the resistor R during the positive excursion of the pulse and the rate of charging will be comparatively slow. The discharge will be much more rapid, since now the cathode of the diode will be carried negative with respect to the anode and the time-constant for discharge has the

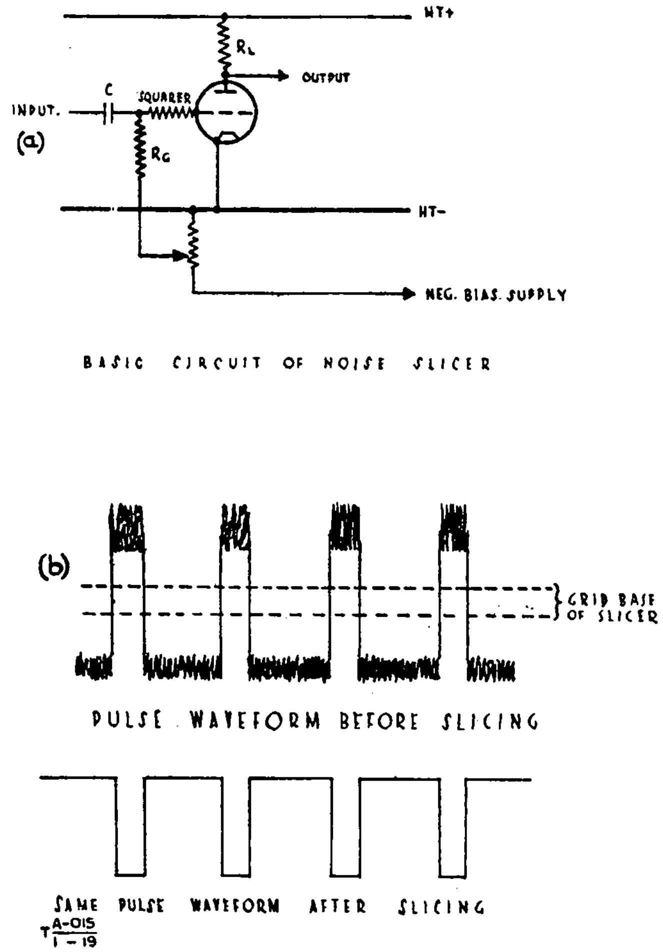


Fig. 19—Application of slicing circuit

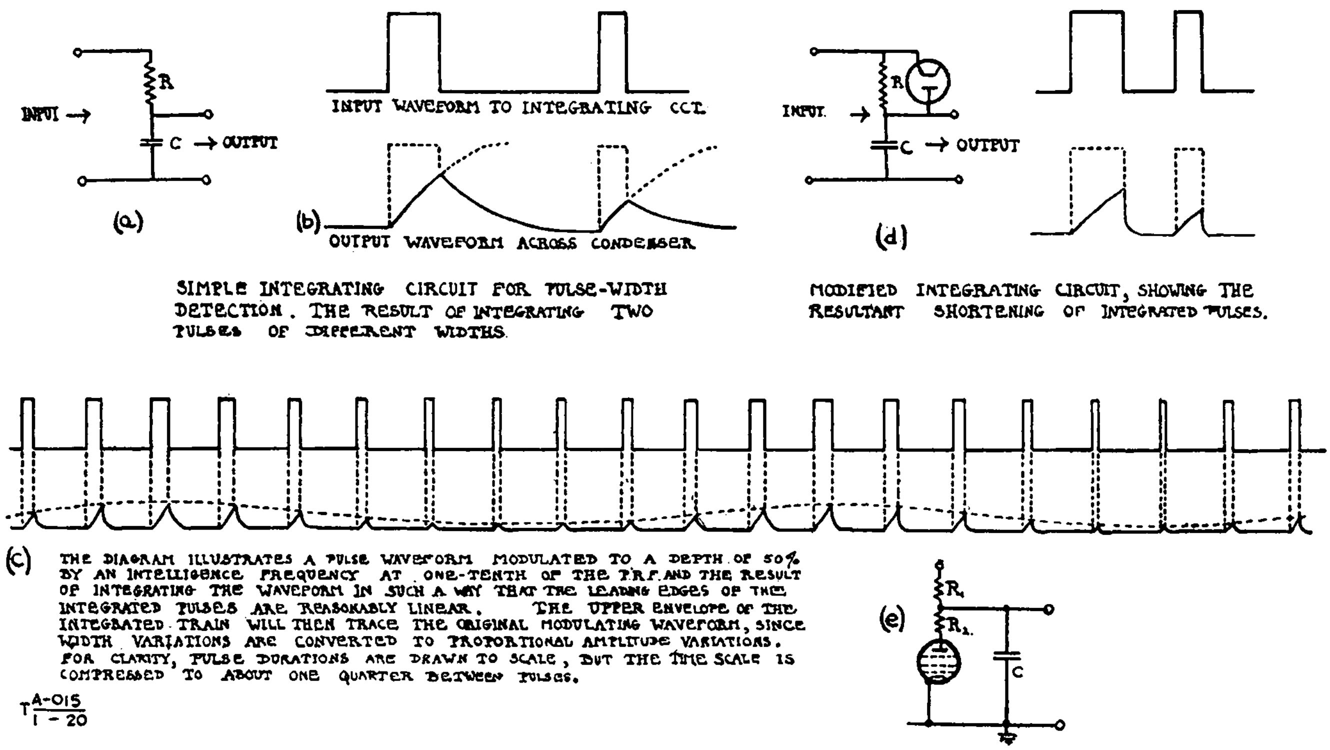


Fig. 20—Conversion of width-modulation to amplitude modulation

impedance of the diode, small compared with R, as its resistance component.

59. Where, as in multi-channel systems, a gating process has to be employed, the width-modulated pulses will usually appear at the anode of the gate valve as negative-going pulses. Integration may then conveniently be carried out in the anode circuit, and such a circuit is shown in schematic form in Fig. 20 (*c*). On the arrival of the front edge of the gated pulse, the condenser seeks to discharge towards the potential reached by the anode of the valve when fully conducting, but the time-constant is so chosen that the voltage excursion is limited to a linear portion of the discharge curve, the integrated pulse so formed having therefore a linear front. The condenser commences to recharge towards the H.T. potential on the arrival of the back edge of the pulse, and since the time interval between pulses is usually very much greater than the pulse duration, the charging excursion is completed before the arrival of the next pulse.

MULTI-CHANNEL COMMUNICATION SYSTEMS EMPLOYING P.W.M.

60. In any wireless system in which a number of channels of duplex voice frequency communication are to be made simultaneously available over a single link, some electrical means of discrimination between the various channels must be employed, since the original frequencies of the intelligence spectra of all the channels initially extend over t same range. Two methods of discrimination are in use at the present time. The first, in which the discrimination is on a *frequency* basis, will be briefly described, for purposes of comparison with the second method in which the discrimination is on a *time* basis, and which will be dealt with in greater detail.

61. In the first method, the intelligence of each channel is first applied to its own channel oscillator, which it modulates in amplitide, with the consequent production in each case of a range of frequencies extending from the central carrier frequency upwards and downwards by the extent of the voice frequency spectrum employed, usually about 3 to 3½kc/s. It is usual to remove the carrier and upper sidebands in each case, leaving only the lower sideband of each channel, as a result of this preliminary modulation. The frequencies of the channel oscillators are so chosen that, after selection of the appropriate sideband in each case, these sidebands do not at any point overlap in frequency, the whole group of frequencies now constituting a spectrum in which the intelligence of each channel occupies its own particular span, distinct from that of any other channel. The sidebands are combined to form one continuum by feeding each to a separate primary winding of a transformer, and taking the output of a single secondary winding. The single group of frequencies so formed is itself now used to modulate a radio-frequency transmitter, and the resultant R.F. frequency spectrum, consisting of the R.F. carrier with corresponding upper and lower sidebands, is radiated. At the receiver it is subjected to the normal processes of frequency changing and amplification and then to first detection. The last process extracts the envelope of the I.F. wave form and so produces an output which contains all the original intelligence, but now still spread out in the continuance spectrum form as after the preliminary modulation. The whole output is applied, in each channel, to a band-pass filter which selects that particular part of the spectrum appropriate to that channel, and the resulting restricted band of frequencies is applied to a demodulator to which also is fed the original channel oscillator frequency. The output of the demodulator consists of the original intelligence, which has thus been transferred over the wireless link without interference from other channels. Nevertheless, in order to avoid intermodulation or crosstalk between channels, even after frequency translation has been carried out by the preliminary modulating processes, it is necessary to ensure complete linearity throughout the subsequent circuits up to the final demodulators, and this is somewhat difficult to achieve in practice. The privacy of the system may be very good indeed, provided that there is no "carrier leak" from the original modulating processes. The range of such a wireless link is not in practice likely to be much greater than that of an equivalent P.W.M. system, now to be described, employing a similar number of channels, since the large band-width requirements of the former necessitates the use of V.H. or U.H. radio frequencies, with consequent limitation of path to optical. Fig. 21 illustrates the principles involved.

62. In the second multi-channel method, discrimination between the communication channels is on a time basis. In Fig. 22 (*a*) two elements of a pulse train are shown on an extended time scale, and it should be noted that the pulse width is small compared with the time interval between the pulses. Provided that the pulse recurrence frequency is at least three times the highest modulation frequency to be transmitted, the pulse train will be capable of conveying intelligence by normal width modulation.

63. Now it is possible to generate and propagate another system of pulses, having the same recurrence frequency, and width-modulated by the intelligence corresponding to a second channel of communication but displaced along the time axis so that they do not anywhere overlap in time the pulses of the first channel, shown in Fig. 22 (*a*). Such a second system of pulses is shown in Fig. 22 (*b*), the same time scale being used here as in the first figure. Extending the idea still further, Figs. 22 (*c*) and (*d*) show two further trains of pulses plotted on the same time scale, and none of the four trains now drawn can interfere in any way with any of the other three. Provided that the generating and modulating processes are separate and distinct for each of the four trains, and provided that some method of selection or "gating" the pulses appropriate to each channel of communication is available at the receiving end of the wireless link, then the separate trains of pulses may be combined, as in Fig. 22 (*e*) and used to modulate a single radio transmitter, with the production of an R.F. train of similar geometrical aspect.

64. It must clearly be a condition for success in avoiding crosstalk between channels that, for any possible width of pulse possible on any channel, no overlap of pulses may take place on the time axis. Should this condition not be observed, then any gating process used will not allow discrimination between the pulses of the two overlapping channels. Further, to facilitate setting-up of the gating circuits and allow the necessary latitude for operational contingencies, it is usual to allow a certain margin over and above the least time theoretically required between successive pulses in the combined train.

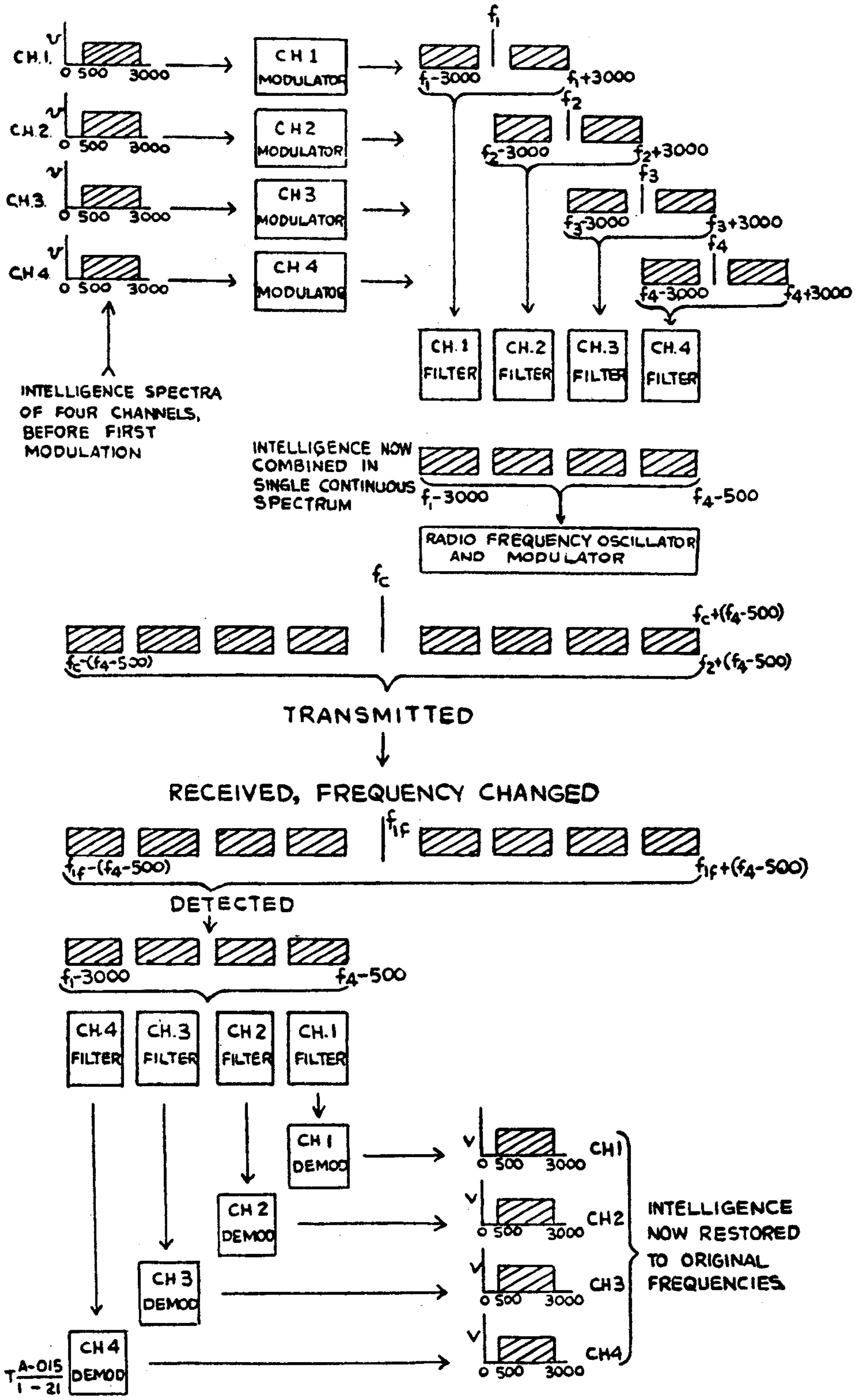


Fig. 21—Method of frequency discrimination

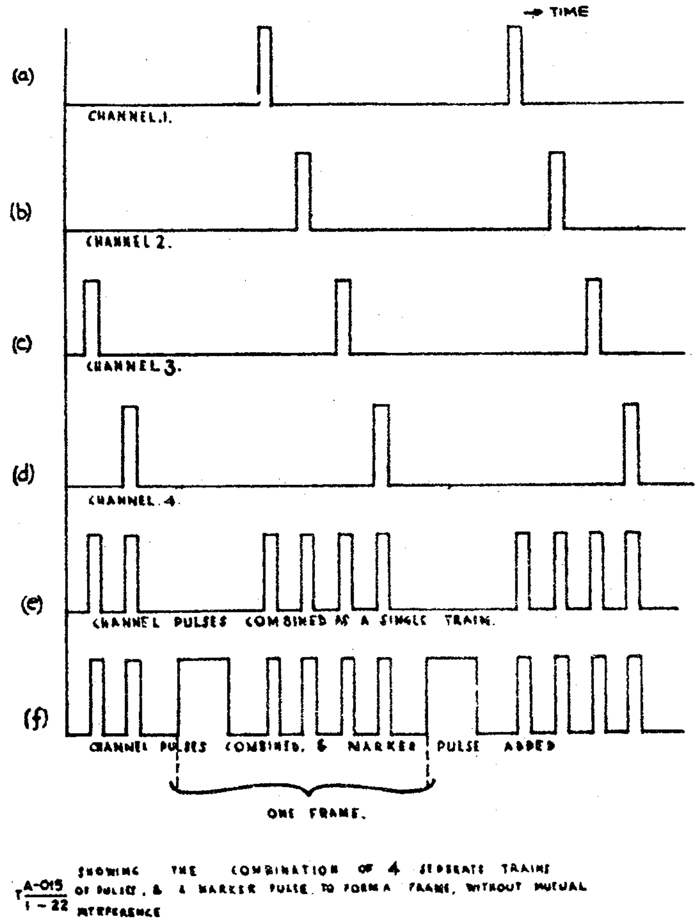


Fig. 22—Method of time discrimination

65. The limit to the number of channels which may usefully be employed on such a system depends mainly on two factors, namely, pulse width and pulse recurrence frequency. The narrower the mean pulse width used for a given pulse recurrence frequency, the more pulses of other channels that can be interposed in time between successive pulses of a single channel. As has been noted in the main text, continuous reduction of the pulse width in these circumstances will also reduce the wide band gain of the system, and it is thus this factor which imposes the final theoretical limitation on the number of channels. As regards the pulse recurrence frequency, the lower this is, the greater will be the time interval between the successive pulses on any one channel, and therefore, for a given pulse width, the greater the number of channels that can be accommodated. The lower limit of pulse recurrence frequency has been shown to the about three times that of the highesr modulating frequency to be transmitted, which thus sets the upper limit to the number of channels possible once the pulse mean width has been chosen.

66. This method of interlacing the pulses of different channels of communication on the time axis will require some form of timing signal so that the gating sequence at the receiver may be carried out successfuly, and the pulses appropriate to each channel selected and passed to the width detector circuits. The timing of the gating is achieved by fitting into the pulse train at the transmitter a synchronizing or "marker" pulse of such width that it may be easily distinguished by a "sync.-separation" circuit at the receiver, and used to time the generation of the gate pulses required in each channel. The whole wave form, from the onset of one synchronizing pulse to the onset of the next, is known as a "frame" of pulses, and the frame recurrence frequency is thus the same as the pulse recurrence frequency for any one channel. A typical frame of D.C. pulses for a four-channel system is shown in Fig. 22 (*f*). The marker pulse is seen to be about 4–6 times the width of a single channel pulse, and this is of the order of width difference usually employed.

67. The gating sequence referred to above is carried out in the following manner. The D.C. pulse train which results, at the radio receiver, from the process of first detection followed by slicing, is passed to a number of gate valves, one per channel of communication. These valves are normally biased off completely, as far as flow of anode current is concerned, by negative voltages at *two grids*, perhaps the control grid and the suppressor grid in the case of a pentode. The whole of the incoming wave form is fed to one of the grids, while to the other grid is fed a positive gate pulse of sufficient width to accommodate any given channel pulse but not wide enough to cover two pulses. Thus, in order that any flow of anode current may occur, both a gate pulse and a channel pulse must be appearing at their respective grids simultaneously, when current will flow for the time for which they overlap. Thus, in order that a replica of the channel pulse may appear in a resistive anode load, the gate pulse must arrive slightly before, and be completed slightly after, the channel pulse which it is gating. The gate pulses for each channel must therefore be adjusted to the correct width, and be timed to appear at the correct moment to gate the appropriate pulse on each frame in each channel.

68. The timing of the gate pulses appropriate to each channel is controlled by the marker pulse on each frame. The marker pulse is separated from the channel pulses by passing the whole pulse train to sync.-separation circuits, which may consist of a simple integrating circuit, converting the large width difference between the synchronizing pulse and any of the channel pulses to a corresponding voltage difference, followed by a discriminator circuit, which utilizes this voltage difference to separate completely the synchronizing pulse from the channel pulses. This separated synchronizing pulse is now used to trip off a circuit which generates, in the correct sequence, gate pulses in each of the channel gate pulse generator stages. The gate pulses so formed are fed to the gate valves with results indicated above. The gate pulse generators will have provision for adjusting both the width and the time positions of the gate pulses produced.

69. The output of each gate valve will consist of the width-modulated train of channel pulses appropriate to that particular channel. This train will normally now be fed to a width detector circuit, followed by a low-pass filter, when the intelligence will finally be reproduced.

The basic requirements of a P.W.M. multi-channel system

70. These are at the transmitting station :—

(*a*) An oscillator, operating at the frame recurrence frequency, and used to time the generation of the synchronizing pulse and all channel pulses on each frame.

(*b*) A synchronizing pulse generator, triggered by the above oscillator.

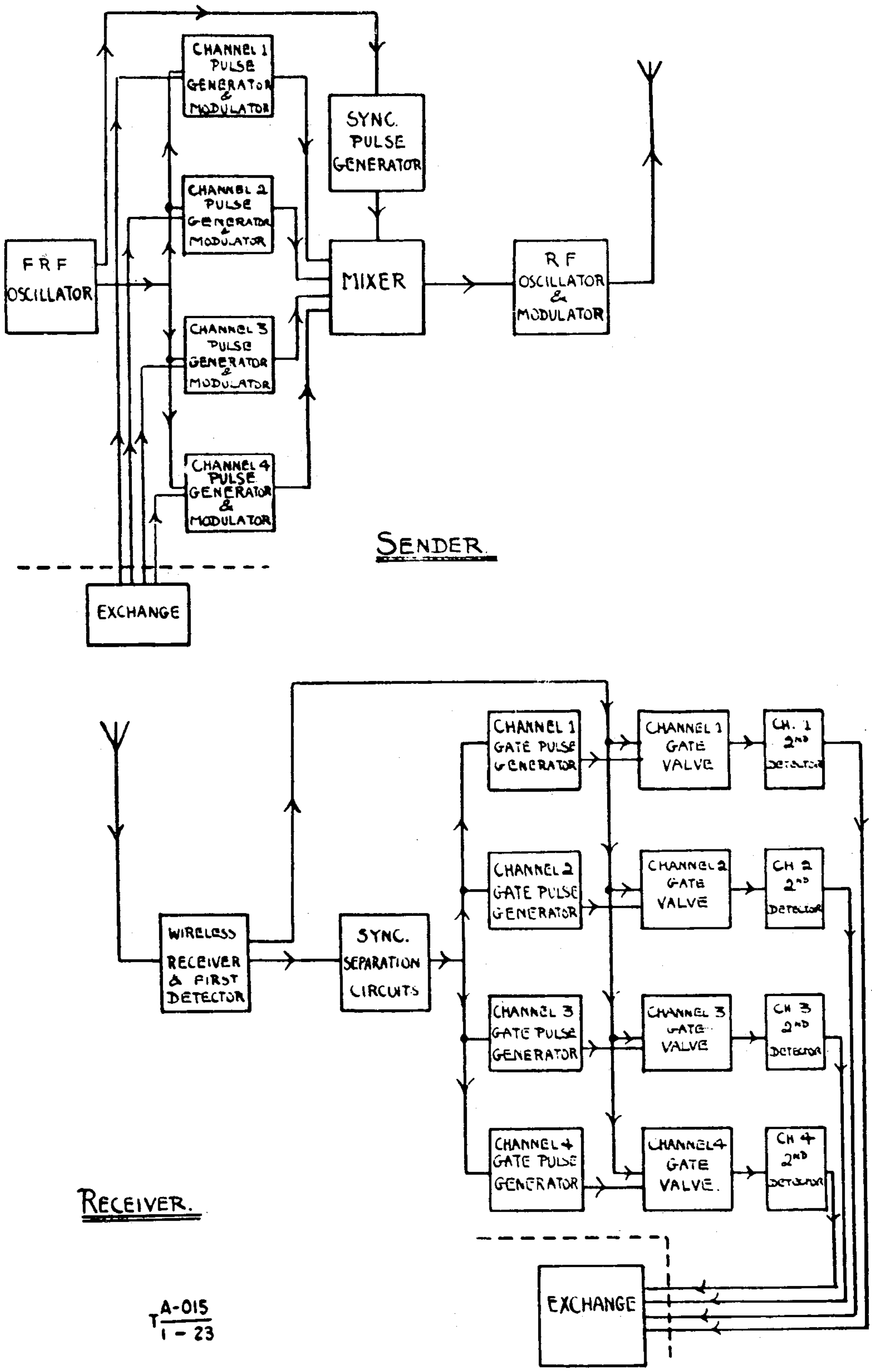


Fig. 23—Essentials of a 4-channel pulse-width-modulated system

(c) One channel pulse generator for each channel of communication, triggered by the above oscillator, combined with a modulator stage which allows the channel pulses to be width-modulated by the intelligence.
(d) A video-frequency mixer, to receive, and combine in a single wave form, all the pulses on each frame.
(e) A radio-frequency oscillator, and means of modulating it with the D.C. output of the video mixer.

71. At the receiver, the requirements will be :—
(a) Superhet. receiver and first detector, followed by noise slicer.
(b) Sync.-separation stages.
(c) One gate pulse generator per channel, triggered with appropriate delay by a signal derived from the synchronizing pulse of each frame.
(d) One gate valve per channel, fed with the whole of the frame of D.C. pulses, and one gate pulse per frame.
(e) One width-modulation detection circuit and filter per channel.

A complete block schematic of a 4-channel system is shown in Fig. 23.

72. It is now possible to compare the two types of multi-channel systems described above. In both of them two processes of modulation and two of detection are required. In both of them increasing the number of channels is likely to raise the band-width requirements of the receiver for different reasons, and so result in poorer comparative signal-to-noise ratios. The methods of separation of the channels at the receiver are analogous in the two cases, if we may regard the gate valve as a type of "time-filter" corresponding to the frequency band filter of the frequency discrimination method. The radio frequency requirements of the two methods are likely to be of the same order, except that, as we have seen, centimetre oscillators become available for the P.W.M. system. Operational ranges, where the number of channels employed is comparatively large, would tend towards optical in both systems, because of the band-width requirements. Power requirements of the transmitter are likely to be very much in favour of the pulse transmission system, for reasons which have been discussed in the main text, and this is borne out in practice, where, for example, we have the W.S.10 giving eight channels of communication over optical distances, and requiring a mean power of about ¼W, as opposed to the W.S.26, which gives six channels of communication over approximately the same distances, and requires between 50 and 100W. In complexity of design there is perhaps little to choose between the two methods, but the possibility of using centimetre oscillators of low power, with consequent simplicity in aerial design tends to make the weight of the pulse transmission system very much less than that of the A.M. system. The comparison therefore seems to show that the pulse system will prove a better practical proposition than the A.M. system.

73. There is one further very important aspect not yet taken into account in the comparison, and that is the question of crosstalk between channels. In the time discrimination method, it is theoretically possible, and possible in practice too, completely to prevent crosstalk between channels by employing the gating process and ensuring that the circuits through which the pulses pass have time for complete recovery after the passage of one pulse before the appearance of the next, a condition which is secured easily by choosing the correct spacing of the pulses in each frame. In the frequency discrimination method such lack of interference between the channels is possible only provided that the circuits after the preliminary modulator and up to the final demodulator are quite linear in their response, since all the intelligence is transmitted simultaneously. Such linearity is somewhat difficult to achieve in practice.

MATHEMATICAL TREATMENT OF PULSE-WIDTH MODULATION

74. Four aspects are considered here :—
(a) The band-width required by a P.W.M. system.
(b) The actual frequency content of a width-modulated train of pulses.
(c) The method of width modulation
(d) Comparison of A.M. and P.W.M. signal-to-noise performance.

Band-width requirements of P.W.M. system

75. An analytical approach to this problem has been devised, but it is not proposed to include it in this discussion. It does, however, appear to support a simple argument based on rather more physical grounds, which is as follows. Let us suppose that the R.F. and I.F. band-width provided by a P.W.M. receiver is sufficient to pass a train of pulses, each of the shortest duration which occurs in the actual modulated train, without reduction in amplitude. The condition for this has already been given in the text, and we may quote it in the form $B = 2/a_{min.}$, where B is the band-width in Mc/s and where $a_{min.}$ is the shortest pulse duration in microseconds. Then it seems reasonable to expect that the receiver will handle equally well a width-modulated pulse train in which $a_{min.}$ is the width of the narrowest pulse occurring, especially if the separation of the successive pulses is large compared with their individual widths. Thus, in this case, we may calculate our required band-width by consideration of the least pulse width.

76. In a multi-channel system, the interlacing of the trains of channel pulses has two effects : it reduces the spacing between successive pulses, and also introduces arbitrary and disconnected changes in pulse width from one pulse to the next. In these circumstances we can expect our previous result to hold only if the response of each pulse is in no way "conditioned" by the previous pulse, i.e., as long as the receiver circuits have time to return to a completely inert state before each pulse arrives. In the practical case of the W.S.10 equipment the least separation between pulses is approximately 5μsec., and this should be adequate, as the response of each pulse will have collapsed almost to zero in $2/B\mu$sec., which in this case, with $B=4$ (Mc/s), is equal to $\frac{1}{2}\mu$sec.

Frequency content of a width-modulated pulse train

77. We establish first of all a useful subsidiary theorem. Let $f(t, \alpha)$ be a periodic function of t, of period T_r, where T_r is independent of α. Suppose also that $f(t, \alpha)$ can be

expanded in a Fourier series of all values of α in a certain range. Then we may write :

$$f(t, \alpha) = A_0(\alpha) + \sum_{n=1}^{\infty} \left\{ A_n(\alpha) \cos \frac{2\pi nt}{T_r} + B_n(\alpha) \sin \frac{2\pi nt}{T_r} \right\}$$

78. Now suppose that α varies in any manner whatever with t, while remaining within the range of values over which the above expression is valid, say,

$$\alpha = \varnothing(t)$$

Then we shall show that

$$f\left\{t, \varnothing(t)\right\} = A_0\left\{\varnothing(t)\right\} + \sum_{n=1}^{\infty} \left[A_n\left\{\varnothing(t)\right\} \cos \frac{2\pi nt}{T_r} + B_n\left\{\varnothing(t)\right\} \sin \frac{2\pi nt}{T_r} \right] \quad (1)$$

This might appear to be obvious from the previous statement, but against this it may be pointed out that the R.H.S. is no longer a Fourier expansion, because its terms are no longer necessarily periodic in t, ; we may describe it as a quasi-Fourier expansion. However, a few simple steps will establish fully the correctness of the equation quoted.

79. Consider any instant $t = t_1$. At this instant the R.H.S. of the equation has the value :—

$$A_0\left\{\varnothing(t_1)\right\} + \sum_{n=1}^{\infty} \left[A_n\left\{\varnothing(t_1)\right\} \cos \frac{2\pi nt_1}{T_r} + B_n\left\{\varnothing(t_1)\right\} \sin \frac{2\pi nt_1}{T_r} \right]$$

which is the value at $t = t_1$ of the series

$$A_0\left\{\varnothing(t_1)\right\} + \sum_{n=1}^{\infty} \left[A_n\left\{\varnothing(t_1)\right\} \cos \frac{2\pi nt}{T_r} + B_n\left\{\varnothing(t_1)\right\} \sin \frac{2\pi nt}{T_r} \right]$$

This last series is actually a Fourier expansion and for all values of t converges to a value of

$$f\left\{t, \varnothing(t_1)\right\} \quad \ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \quad \ldots\ldots (2)$$

except for values of t at which $f(t, \alpha)$ is not continuous. Thus the R.H.S. of equation (1) is equal to the value of the equation (2) at time $t = t_1$ which is :—

$$f\left\{t_1, \varnothing(t_1)\right\}$$

But this is true whatever the value of t_1, apart from the points of discontinuity already mentioned, so that the validity of equation (1) is established. For full analytical rigour the behaviour of the series at the points of discontinuity should be considered, but it is sufficient to state here that in any practical case we may consider, where the discontinuities are finite, they do not give rise to any significant change in the argument.

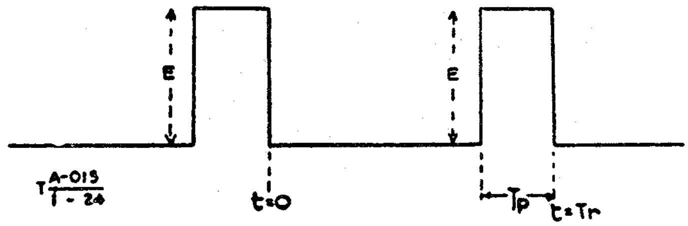


Fig. 24—Train of rectangular pulses

80. Let us consider the Fourier analysis of a train of rectangular pulses of height E, width T_p, and recurrence period T_r as shown in Fig. 24. Then the function to be analysed is defined by :—

$$f(t) = 0, \quad 0 < t \leqslant T_r - T_p$$
$$f(t) = E, \quad T_r - T_p < t \leqslant T_r$$

$$\text{Let } f(t) = a_0 + \left\{ a_1 \cos \frac{2\pi t}{T_r} + b_1 \sin \frac{2\pi t}{T_r} \right\} + \left\{ a_2 \cos \frac{2\pi .2t}{T_r} + b_2 \sin \frac{2\pi .2t}{T_r} \right\} + \ldots\ldots + \left\{ a_n \cos \frac{2\pi nt}{T_r} + b_n \sin \frac{2\pi .nt}{T_r} \right\} + \text{ad inf.} \ldots\ldots (3)$$

If we integrate both sides from 0 to T_r, we get

$$\int_0^{T_r} f(t)\,dt = a_0 T_r$$

since all the other integrals on the R.H.S. vanish. This reduces to

$$E.\,T_p = a_0 T_r$$

$$a_0 = E\,\frac{T_p}{T_r}$$

Multiplying by $\cos \frac{2\pi nt}{T_r}$ on each side of equation (3), and integrating from 0 to T_r, we get :—

$$\int_{T_r - T_p}^{T_r} E \cos \frac{2\pi nt}{T_r}\,dt = \frac{T_r\,a_n}{2}$$

$$\text{or } \frac{T_r . a_n}{2} = \left[E \sin \frac{2\pi nt}{T_r} \cdot \frac{T_r}{2\pi n} \right]_{T_r - T_p}^{T_r}$$

$$\text{or } a_n = -\frac{E}{\pi n} \sin \frac{2\pi n (T_r - T_p)}{T_r}$$

$$= \frac{E}{\pi n} \sin \frac{2\pi n\,T_p}{T_r}$$

By multiplying through by $\sin \frac{2\pi nt}{T_r}$ we obtain similarly :—

$$b_n = \frac{E}{\pi n}\left\{\cos\frac{2\pi n T_p}{T_r} - 1\right\}$$

Thus, substituting back in equation (3), we get :—

$$f(t) = \frac{E}{\pi}\left[\frac{\pi T_p}{T_r} + \sum_{n=1}^{\infty}\frac{1}{n}\left\{\sin\frac{2\pi n T_p}{T_r}\cos\frac{2\pi nt}{T_r} - \left(1-\cos\frac{2\pi n T_p}{T_r}\right)\sin\frac{2\pi nt}{T_r}\right\}\right] \quad \ldots\ldots (4)$$

Writing $\frac{T_p}{T_r} = \frac{\alpha}{2\pi}$, we get :—

$$f(t) = \frac{E}{\pi}\left[\frac{\alpha}{2} + \sum_{n=1}^{\infty}\frac{1}{n}\left\{\sin n\alpha \cos\frac{2\pi nt}{T_r} - (1-\cos n\alpha)\sin\frac{2\pi nt}{T_r}\right\}\right] \quad \ldots\ldots (4a)$$

$$= \frac{E}{\pi}\left[\frac{\alpha}{2} + \sum_{n=1}^{\infty}\frac{1}{n}\left\{-\sin\frac{2\pi nt}{T_r} + \sin\left(\frac{2\pi nt}{T_r} + n\alpha\right)\right\}\right] \quad \ldots\ldots (4b)$$

Thus the train of pulses considered can be analysed into a steady component together with a series of sinusoidal components whose frequencies are $\frac{1}{T_r}$ ($= f_r$, say) and its harmonics.

81. We may now apply the subsidiary theorem established at the beginning of this section. Suppose that α, instead of being constant, suffers a sinusoidal modulation in amplitude α_1, and angular frequency ω_m ($=2\pi f_m$) about a mean value α_0, say,

$$\alpha = \alpha_0 + \alpha_1 \sin \omega_m t \quad \ldots\ldots (5)$$

where α_1 is less than α_0. We may also suppose that this change in α, or T_p is produced linearly by some modulating voltage. Then the first term in the expansion 4(b) for f(t) becomes

$$\frac{E}{2\pi}\left\{\alpha_0 + \alpha_1 \sin \omega_m t\right\}$$

so that the modulated pulse wave form contains a component proportional to the modulating voltage. We may allow ourselves a slight digression here to see how this connects with the considerations put forward on the question of signal-to-noise ratio. In the W.S.10 equipment, to quote a practical case :—

$$T_p = 3\tfrac{1}{3} \pm 2\tfrac{2}{3}\ \mu\text{sec.} \quad T_r = \frac{1000}{9}\ \mu\text{sec.}$$

Hence $\alpha_1 = \frac{17/6}{1009/9} = .025$ at most.

Thus the modulation frequency component which can be extracted from the modulated pulse train is only about 1/40 of the pulse height in amplitude, whereas, *in the absence of a slicer*, the noise amplitude might be, under the conditions postulated for comparison with A.M., considerably greater than this. Thus it appears that, but for the action of the slicer, the S/N ratio would compare very unfavourably with the A.M. case.

82. Reverting to equation 4(b), we may now consider the effect of the variation in α on the remaining terms. The first term under the Σ is constant with respect to α, so we have only to consider the other, and this becomes

$$\frac{E}{\pi n}\sin\left(\frac{2\pi nt}{T_r} + n\alpha_0 + n\alpha_1 \sin \omega_m t\right)$$

This is of the form of a phase-modulated sinusoid, of centre frequency $\frac{n}{T_r} = nf_r$ with maximum phase deviation $n\alpha_1$. By P.M. theory, this splits up into sinusoidal components as follows :—

(*a*) A centre frequency component of amplitude

$$\frac{E}{\pi n}\left| J_0 (n\alpha_1) \right|$$

(*b*) Side frequency components $nf_r \pm sf_m$ ($s = 1, 2, 3$, etc.) each of amplitude

$$\frac{E}{\pi n}\left| J_s (n\alpha_1) \right|$$

where the Js are Bessel functions. It may be mentioned in passing that it is this analysis of the modulated pulse train which can be used as a starting point for an attack on the question of band-width.

83. Thus frequencies f_m, $nf_r \pm sf_m$ ($n = 1, 2, 3$, etc.) are present in the modulated pulse train; and if the modulation contains frequencies f_{m1}, f_{m2}, etc., then by an argument similar to the above we can show that frequencies of the form $nf_r \pm s_1 f_{m1} \pm s_2 f_{m2}$ ($s_1, s_2 = 0, 1, 2, 3$, etc.) will also be present. Thus a width-modulated pulse train contains a complex group of frequencies, and it will be appreciated that it is not very easy to set up a satisfactory criterion for the band-width sufficient to ensure good reproduction of the whole pulse train.

84. Another important phenomenon we may note here arises from the above. The term in the spectrum of a pulse train carrying a purely sinusoidal modulation, of frequency corresponding to that modulation frequency, arises from the first term in the original Fourier expansion. But frequencies such as $f_r - f_m$, $f_r - 2f_m$ etc., are also present, and some of these may fall within the A.F. range being used. As there is no means of removing them, they consitute a source of distortion, and if, as is sometimes done, a number of P.W.M. wireless links are used "in series," this distortion will be cumulative. The severity of the distortion depends on the amplitudes of the spurious A.F. components, which in turn are determined by the values of the appropriate Bessel functions. Thus, for

example, if we take $n = 1$, $s = 2$, $f_r = 9$kc/s, $f_m = 3$kc/s, then the component of the above type of frequency $f_r - 2f_m$ (i.e. 3kc/s) will have amplitude

$$\frac{E}{\pi} \left| J_2(\alpha_1) \right|$$

where α_1 may lie between 0 and $0.02 \times 2\pi$. This gives a maximum value for the amplitude of about

$$\frac{E}{\pi} \times .002$$

and we may compare this with the amplitude of the 3kc/s component arising from the first term of the expansion, namely,

$$\frac{E}{\pi} \times \frac{\alpha_1}{2} \doteqdot \frac{E}{\pi} \times 0.066$$

Thus the distortion in any one link is apparently slight, although its effect may be aurally unpleasant when the frequencies concerned do not enjoy the above simple relationship.

The method of width modulation

85. We may write equation (5) of the previous section in the more general form :—

$$\alpha = \alpha_0 + \rho v$$

where v is the instantaneous value of the modulating voltage, not necessarily sinusoidal, and ρ is a constant. Suppose we measure time t^1 from the instant of the onset of the pulse when unmodulated, and express v as a function $v(t^1)$ of this new variable. Suppose also that the pulse starts, when modulated, at time $t^1 = t_1$. Then

$$\alpha - \alpha_0 = -\frac{2\pi t_1}{T_r}$$

so that t_1 is determined by the equation :—

$$-\frac{2\pi t_1}{T_r} = \rho v(t_1)$$

the solution of which can be determined by finding the point of intersection of the graphs

$$y = -\frac{2\pi}{\rho T_r} t^1 \text{ and } y = v(t^1)$$

It will be seen from Fig. 25 that this involves exactly the same state of affairs as has been previously envisaged in describing the type of modulation employed in practice, so that paras. 77 to 84 are concerned with modulation of this type. The result of paras. 77 to 84, that the width-modulated pulse train contains a component proportional to the modulation voltage, thus applies to this type of modulation, the only restriction being that the distortion components discussed in para. 84 shall be negligible.

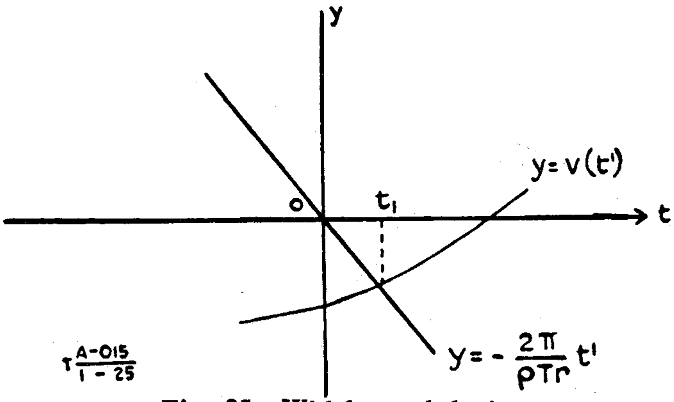


Fig. 25—Width modulation

Comparison of the A.M. and P.W.M. signal-to-noise performance

86. The following is an attempt to make the above comparison on simple grounds, with a statement of each assumption made and some estimate of its validity. Any disagreement with published researches on the subject will probably be found to arise from variation in these assumptions. The comparison involves three systems, which are as follows :—

(*a*) A P.W.M. system using a mean pulse width *a*, with a maximum deviation from this value of *d*, leading-edge modulation being used. The pulse recurrence frequency is f_r, and the highest modulating frequency f_m. The P.W.M. receiver has its band-width B c/s, large enough to allow it to pass the narrowest pulse without reduction in amplitude. The receiver is provided with a slicer, and a low-pass filter with uniform response up to its cut-off frequency, which is f_m.

(*b*) An A.M. system, in which the receiver has a band-width of B c/s; and the maximum modulating frequency is f_m as before.

(*c*) An A.M. system identical with (*b*) except that the receiver has a band-width of $2f_m$.

87. We assume that all the band-widths are idealized, i.e., response is constant within them and zero outside, with no phase distortion. This gives a good approximation to the action of practical circuits, even though the characteristics just quoted cannot be realized in practice.

88. Of the three systems mentioned in para. 85, (*a*) and (*c*) are the ones we actually wish to compare ; (*b*) is an intermediate case linking the other two. We take it that the mean signal power fed into the three receivers is the same, and also that the noise power generated in each is the same per unit of band-width, so that the total noise power is the same in (*a*) and (*b*), but smaller in (*c*). The term S/N ratio will be used only for the signal-to-noise ratio in the A.F. output.

89. Suppose that in system (*b*) an unmodulated signal power P is sufficient to give a carrier of amplitude twice that of the noise. Then the noise will give rise to A.M. of depth 50%, as against a possible 100% signal modulation. As the latter requires a 50% increase of signal power, we obtain the result that with input signal power 3P/2 an optimum S/N ratio of 2 is obtained, measured in terms of *voltage*. Now in system (*c*) the same signal input power will give rise to the same A.F. voltage output, but the noise power generated will be reduced in the ratio $2f_m$: B. Thus the noise voltage output will be multiplied by a factor $\left[\frac{2f_m}{B}\right]^{\frac{1}{2}}$ and for this system the S/N ratio becomes $2 \times \left[\frac{B}{2f_m}\right]^{\frac{1}{2}}$. Finally, in system (a), let us assume that in the first place a signal of pulse power P^1 ($= kP$, where *k* is greater than unity) is fed into the receiver. Then in the output of the first detector the pulse com-

ıonent will have an amplitude of $2k^{\frac{1}{2}}$ times greater than ˑhat of the noise component. Thus the condition of Fig. 3 (*c*) exists and operation of the slicer is possible.

0. It has already been pointed out in para. 31 that if the ulse component of the first detector could be perfectly ectangular, then no noise could appear on the sides of the ulses, or, as a result, in the A.F. output. Thus the ıtrusion of noise is due to degeneration of the pulse shape :sulting from restriction in band-width. We shall accept ıe result already used that if the rectangular pulse is fed ıto a network of band-width B c/s, then the output pulse ·ill be trapezoidal, with sloping sides, each of duration 'B sec. This result can be argued quite closely from the :revious ones made about the ideal performance of a etwork within its band-width. Hence, in the present ıse, the pulse component of the output of the first de- ctor will be of the form just quoted. The slicer can be :t to operate at two levels close together, of which the wer bisects the sloping sides of each pulse; as a result ıe slicer output would be, in the absence of noise, of the me duration as the original rectangular pulse; more- ·er, except during the sloping sides of each pulse, the icer levels will be clear of noise. We have now to con- ler the noise that will arise in the A.F. output as a result the noise on the sloping sides of the pulse. Fig. 26 a diagram of the leading edge of the pulse.

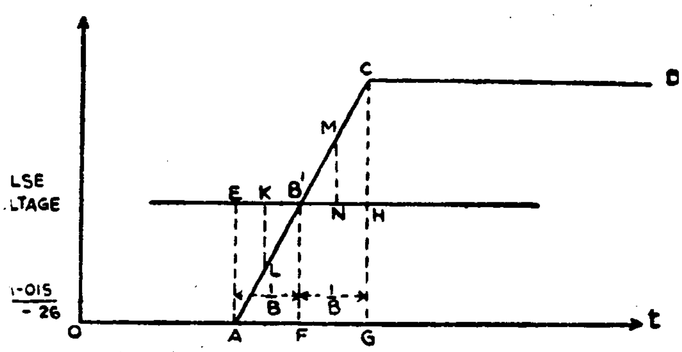


Fig. 26—Leading edge of pulse

In this diagram the time interval $KB' = 1/Bk^{\frac{1}{2}}$, that $KL =$ pulse amplitude$/2k^{\frac{1}{2}}$ = noise amplitude. ıs when the pulse wave form has reached the point and not before, it may happen that the instantaneous ıe of the noise voltage present is sufficient to lift the l voltage to that value required for operation of the :r, so advancing the start of the pulse by $1/Bk^{\frac{1}{2}}$ sec. ılarly the start of the pulse may be delayed to N, :re $B'N = KB'$. The trailing edge of the pulse may er the same displacement in either direction, so that pulse as a whole may be lengthened or shortened by ımount $2/Bk^{\frac{1}{2}}$. We shall assume that the net effect ıe same as the variation of this amount operating on the .ing edge only. This assumption ignores the phase ·lacement resulting from the displacement of the trailing : of the pulse; it does in fact seem likely that the sub- ıent demodulating circuit will prove insensitive to this.

Thus we now have a pulse width variation due to :e of amplitude $2/Bk^{\frac{1}{2}}$ which might appear could be ıpared directly with the maximum variation due to ıal modulation, namely, d. There are, however, two ɔrs still to be taken into account :—

(*a*) The mean signal power has still to be equated to that used in the A.M. case.

(*b*) The effect of the A.F. filter has to be considered.

In the P.W.M. system the pulse power kP is present only for a time af_r in each second, so that the mean power is af_rkP, and this is still the correct value during modulation to any depth. Hence we must equate this mean power to that used in the A.M. systems under the conditions in which the optimum signal-to-noise ratio is obtained, i.e., to 3P/2.

This gives :—

$$af_rk = 3/2$$

$$\text{or} \quad k = 3/2af_r$$

and hence the pulse width variation due to noise becomes

$$\frac{2}{B} \times \left[\frac{2af_r}{3}\right]^{\frac{1}{2}}$$

and the optimum signal-to-noise ratio in the absence of an A.F. filter is obtained by dividing d by this expression, which gives :—

$$\frac{dB}{2} \times \left[\frac{3}{2af_r}\right]^{\frac{1}{2}}$$

93. The action of the A.F. filter may now be considered as follows. We assume that the A.F. noise spectrum extends from 0 to f_r with uniform amplitude; this in turn depends partly on the assumption that width modulation of a pulse train of recurrence frequency f_r at a modulating frequency above f_r is impossible. This last is definitely untrue, so that our original assumption ignores part of the noise spectrum, and we are thus assuming a greater concentration of noise energy between 0 and f_r than actually exists, thereby undervaluing the action of the A.F. filter in improving the S/N ratio. Thus on the original assumption the A.F. filter, which passes only frequencies up to f_m, will multiply the noise energy content of the output by a factor f_m/f_r, which is less than unity. This factor would of course have been smaller if we had assumed an extension of the A.F. noise spectrum beyond f_r. Thus the A.F. noise amplitude will be multiplied by a factor $(f_m/f_r)^{\frac{1}{2}}$, and the final expression for the S/N ratio is

$$\frac{dB}{2} \times \left[\frac{3}{2af_r}\right]^{\frac{1}{2}} \times \left[\frac{f_r}{f_m}\right]^{\frac{1}{2}}$$

Hence, using the expression previously obtained for the S/N ratio of the A.M. system (iii), we get :—

$$\text{Wide-band gain of P.W.M.} = \frac{\text{S/N ratio of P.W.M. system}}{\text{S/N ratio of A.M. system}}$$

$$= \frac{\frac{dB}{2} \times \left[\frac{3}{2af_r}\right]^{\frac{1}{2}} \times \left[\frac{f_r}{f_m}\right]^{\frac{1}{2}}}{2 \times \left[\frac{B}{2f_m}\right]^{\frac{1}{2}}}$$

This reduces quite simply to the form

$$\frac{(3aB)^{\frac{1}{2}}}{4} \times \frac{d}{a}$$

94. From this it can be seen that, although a certain minimum band-width is essential so that a pulse of the minimum width a–d can be passed without loss of amplitude; it will in fact be advantageous to exceed this band-width so as to improve the wide-band gain. It can also be seen that as long as B can be made large enough, it is also advantageous to make d as nearly as possible equal to a.

95. For the W.S.10 equipment, the above formula gives a value

$$\frac{(3 \times 3.33 \times 10^{-6} \times 4 \times 10^{6})^{\frac{1}{2}}}{4} \times \frac{3.33-0.5}{3.33}$$

$$= 1.34,$$

the mean pulse width employed being 3.33μsec., and the receiver band-width 4Mc/s. The improvement in the signal-to-noise ratio is *one* of the factors which makes it possible to operate the W.S.10 with a smaller mean power than that of an equivalent A.M. system.

D.M.E. Encl. No. 9 to 57/Maintenance/1310.

END